Gizapedia

gizapedia.hirusta.io

STATISTICS FOR BUSINESS

Workbook

Author: Josemari Sarasola

1 Probability calculus

1.1 Combinatorics

1. How many ways are there to complete 3 tasks one after the other? Make the list of those ways. Idem with 4, 5 and 10 tasks.

2. How many ways are there to schedule 4 work days and 2 holidays in the next 6 days?

3. There are 6 workers in a shop. You have to choose 2 of them to work on Saturday. In how many ways can you do it?

4. Among 8 people you have to choose the president, the vice president and the secretary. How many ways are there to do it?

5. 6 men work in a factory. 4 tasks have to be completed. If men are able to complete one task or more, how many ways are there to assign the tasks? Remark: Guys are distinguishable.

6. A firm packs delicatessen products into gift boxes. A box must be filled with 2 big products and 2 small products. The customer can choose among 4 big products and 6 small products. How many ways are there to fill a box?

7. You have to choose a commission of 2 students between 10 students. Order is not relevant. Two students refuse to be together in the commission. How many ways are there to choose the commission? Idem if the commission comprises 4 students.

8. To list the reference codes of the product from a catalog you have to choose some letters. If the catalog includes 3200 products, at least with how many letters must you create the reference code? Remark: The alphabet contains 26 letters.

1.2 Laplace's rule

9. You have to deliver 4 packages to 4 customers, but as you have lost their addresses, you deliver the packages randomly. What is the probability of delivering them in the correct way?

10. A postman has to deliver 4 letters. She chooses randomly the delivery order.
    (a) What is the probability of choosing first the closest recipient?
    (b) What is the probability of delivering A letter immediately before B letter?

11. **Urn problem**: In an urn there are 12 faultless and 4 faulty items. You extract 4 items at one time (or without devolution).
    (a) What is the probability of all of them being faultless?
    (b) What is the probability of being 3 faultless items?
    (c) Which is the event with the bigger probability? Why?

12. **Urn problem**: There are 6 women and 8 men in a group. 4 people have to be drawn randomly.
    (a) What is the probability of all of them being women?
    (b) What is the probability of Anne and Mary being drawn together?
    (c) What is the probability of drawing 2 men and 2 women?
    (d) What is the probability of drawing 2 men and 2 women, if we add 6 more men to the main group?

13. We have 6 A, 8 B and 10 C type pieces in a box. We draw at the same time 6 pieces randomly.
    (a) What is the probability of drawing 3 A, 2 B and 1 C type pieces?
    (b) What is the probability of drawing 5 A type pieces?

14. We have received 4 only one night bookings in our hostel for the next week, but we don't know the exact day. What is the probability of all of them being in different days?

15. There are 5 queues in a supermarket. Customers choose randomly their queue. Among 3 customers, what is the probability of being at least 2 of them in A queue?

1.3 Event algebra

16. There are 345 people in a village. Among those, 123 people get bread at Anne's bakery. There are 102 people who get bread at Laura's bakery. Everybody else makes bread at home. Formulate using event algebra symbols these events and calculate their probability:

    (a) A person gets bread at a bakery.

    (b) A person makes bread at home.

17. 60 people are attending a congress. 40 of them speak only Spanish, 20 of them French. What is the probability of two of them, randomly chosen, understand each other.

18. We have asked some people about the film they saw last week (A or B). With the data we have drawn this contingency table:

The film you have seen	A yes	A no	Total
B yes	61	84	145
B no	109	112	221
Total	170	196	366

    What is the probability of a person seeing *at least* one film?

19. Tomorrow the probability of raining is 0.8. The day after tomorrow the probability of raining is 0.6. The probability of raining during the following two days is 0.5.

    (a) What is the probability of raining *at least* one day?

    (b) What is the probability of no raining during the following two days?

20. A product may have A and B failures, with a probability of 0.04 and 0.11 respectively. Both failures appear with a 0.01 probability.

    (a) What is the probability of a product not having any failure?

    (b) What is the probability of a product having only one failure?

21. Let A and B events. Applying event algebra symbols, quote these complex events:

    (a) A and B don't happen at the same time,

    (b) at least one happens,

    (c) only one happens,

    Let A, B and C events. Applying event algebra symbols, quote these complex events:

    (d) only A happens,

    (e) at least one happens,

    (f) the three events happen at the same time,

    (g) A and B happen, but not C

    (h) at least two events happen,

    (i) no event happen.

22. In a urn there are 12 faulty items, and 46 faultless. We draw 4 pieces randomly at the same time. What is the probability of drawing at least 2 faultless items?

23. We have collected these data from a classroom:

Students	Math	Geography	Biology	Language
A	5.6	3.6	4.0	8.2
B	3.2	6.4	5.6	4.6
C	2.2	3.4	3.4	4.0
D	8.2	8.6	9.7	9.2
E	7.2	4.5	6.4	7.0
F	2.2	5.4	6.6	4.3
G	6.8	5.3	3.9	4.1
H	8.2	7.6	4.4	8.3
I	4.2	5.5	7.5	6.8
J	7.6	8.0	3.3	5.4
K	6.2	8.5	3.5	4.8
L	3.6	3.9	6.5	6.9
M	4.2	3.5	4.5	5.2
N	4.2	5.5	2.3	4.0

Apply this code: MP: math passed; LN: language not passed, and so on. Calculate these probabilities:

(a) P(MP and BP);

(b) P(MN or LP), applying inclusion-exclusion rule;

(c) P(MP or GP or BN), applying inclusion-exclusion rule. Which is the student who doesn't gather in that probability?

24. In a bag you have 10 A type pieces, 20 B type eta 30 C type. You draw randomly 2 pieces, which is the probability of all of them being of the same type?

1.4 Conditional probability. Dependence and independence. Multiplication rule.

25. In a urn we have 22 faultless and 7 faulty pieces. We draw randomly 4 pieces, without devolution.

(a) What is the probability of all being faultless?

(b) What is the probability of the first three being faultless and the last one faulty?

(c) What is the probability of drawing 3 faultless and 1 faulty pieces?

(d) What is the probability of drawing 2 faultless and 2 faulty pieces?

26. In a urn we have 12 faultless and 4 faulty pieces. We draw randomly 6 pieces, without devolution (or at the same time). What is the probability of having 2 faulty pieces or less?

27. 200 men eta 300 women live in a village. We draw a sample of 5 people.

(a) What is the probability of drawing the number of men and women that matches the population proportion? Calculate the probability with devolution as well as without devolution.

(b) What is the probability of drawing one man in the sample drawn without devolution? Compare this result with the former one.

(c) What is the probability of drawing at least one man? Calculate the probability with devolution as well as without devolution.

28. In a factory we produce with total independence day by day. We can produce 1, 2, 3 or 4 items with 0.1, 0.2, 0.3 eta 0.4 probabilities respectively. Within five days, calculate the probability of producing:

(a) on the first two days 2 or more, and on the last two days less than 2,

(b) 1 only on each of 2 days,

(c) always producing the same number of items,

(d) always producing 3 or less,

(e) 4 on at least one day.

29. A share price increases with a 0.4 probability, when the day before increased, and with 0.7 probability, when the day before the price decreased.

(a) What is the probability of the share price increasing on the nest first two days and decreasing on the next two days, provided the share price increased yesterday?

(b) Within the next two days, what is the probability of the share price increasing on one day and decreasing on the otherprovided the share price increased yesterday?

30. You have a 4 pieces batch, but you don't know which of them are faulty or faultless. Trying to improve the quality of the batch, you draw pieces one by one: if the piece you draw is faulty, you withdraw that piece and put into the batch 2 faultless pieces; if the piece you draw is faultless, you return it into the batch. You draw 3 pieces. If you have 2 faulty and 2 faultless pieces in the batch, calculate the probabilities of all the possible outcomes, taking the order into account. If the three pieces you have drawn are *oxo*, what should you deduce about the number of faulty pieces in the urn at the beginning?

1.5 Probability trees

31. 600 men and 400 women have signed up for an insurance policy with a company. Among these, 200 had an accident: 150 men and 50 women. What is the probability of a person having an accident? Give the probability directly as well as by means of the law of total probability, taking into account the sex.

32. Possible daily sales at an auto dealer are 1 to 4, with 0.1 to 0.4 probabilities respectively, with total independence. What is the probability of selling 6 or more autos in two days?

33. A student takes two partial exams in a course. The probabilities of passing these exams are 0.5 and 0.6 respectively. Partial exams are saved and in the final exam the student must take only the failed partial exams. The probabilities of passing these final partial exams are 0.7 and 0.8 respectively. What is the probability of finally passing both exams?

34. Because of budget restictions, a company that sells a type of machine can attend only two exhibitions this year: A and B exhibitions, or alternatively, C adn D exhibitions. Here you have the probabilities for selling the possible numbers of machines:

Numbers →	0	1	2	3
A	0.10	0.20	0.30	0.40
B	0.35	0.30	0.20	0.15
C	0.20	0.20	0.30	0.30
D	0.30	0.30	0.20	0.20

The goal is to maximize the probability of selling 3 machines al least. Which exhibitions should the company attend?

35. We have three applicants for a job. Two exams will be hold through the recruitment; the probability for passing the first exam is 0.6, and 0.3 for the second one. What is the probability of remaining 0, 1, 2 and 3 applicants after the exams?

36. Our stock price increases with a 0.4 probability, if it increased the day before; and decreases with a 0.7 probability, under the same circumstances. Through the next four days, what is the probability of the stock price increasing over two days and decreasing over the other two? Remark: last day stock price increased.

1.6 Bayes theorem

37. In a population, the percentage of people being affected by a disease is 15%. In order to diagnosticate the disease we have performed a test, but not perfectly: among people with disease, the test performs well in the 80% of cases; among people with no disease, the test says the person is ill in the 10% of cases. For Lucy, the test was positive. What is the probability of Lucy of being ill? And what if the test was negative?

38. A, B and C machines produced 2000, 5000 and 3000 items respectively last year. A machines produces faulty items with a 5% probability. B and C machines produce them with 10% and 2% probability respectively. A customer made a complaint for a faulty item. What should you say about the machine it comes from?

39. A factory makes boxes of chocolate:

    - the red box contains 4 A type chocolates, 6 from B type eta 10 from C type;
    - the blue box contains 10 A type chocolates, 4 from B type eta 6 from C type;
    - the yellow box contains 8 A type chocolates, 8 from B type eta 4 from C type.

    We sent to a shop 200 red boxes, 100 blue boxes and 300 yellow boxes. The shopkeeper found two A type chocolate wrapping papers outside the batch. So, he thought somebody opened a chocolate box, picked up two chocolates, ate them and left the papers out there. He must find the box missing 2 A type chocolates. Which box type should he begin searching from?

40. In a test each question has 4 possible answers. The probability of knowing the answer, and therefore, giving the correct answer is 0.7. We know that 10% of student leave the question blank, and the rest of them answer it randomly. A student answered correctly. What is the probability of really knowing the answer? How mahy options sholuld we give in each question in order to, supposed the answer is correct, the probability of knowing it, is 0.96? Idem, with a 0.99 probability.

41. Broadly, the customers of a new product may be 10%, 20%, 30% or 40% of the total customers in the market. There is total uncertainty about the probability of those values. A survey took 10 people and among them 6 claimed they would purchase that product. How should we change the a priori probabilites about the percentage of customers?

42. In the customer database of an insurance company, we have these data: 100 policy holder had an accident last year, 200 policy holder didn't have any accident last year, 400 policy holders didn't have any accident last two years, and 500 policy holders didn't have any accident last thrre years. According to our estimations, a person that had accident last year has a 0.22 probability of having another accident next year, and that probability declines by 0.04 by year without accident. We have just received a claim about an accident, but we know nothing about the policy holder. Into which policy holder group should we enter him?

1.7 Introduction to statistical testing

43. We are going to buy 2 second hand items from a given seller. The seller has 20 items, and as he claims, there are at most 4 faulty items among them. We will buy 2 items, but only if he chooses them randomly from the set of 20 items. We purchase the items and we notice that both are faulty. Is it a coincidence or should we conclude that the seller is lying? $\alpha = 5\%$

44. We are going to market four different flavoured vegetal drinks, but previouly we want to test their liking among potential customers. At first, we think all flavours are the same and hence with the same probability to be liked. We test the 4 drinks with 3 people, and all of them tell us that best drink is A. Should we conclude that they are same, as we think, or that A drink is better than the others? $\alpha = 1\%$.

45. Our supplier has claimed: into a batch of 20 pieces we have purchased, there are at most 3 faulty pieces. In order to test that claiming, we take randomly 5 pieces and among them we find 2 faulty pieces. What should we conclude? Significance level: 2%.

2 Random variables

2.1 Discrete random variables

46. We have 6 faulty washing machines in stock, and 12 faultless. We sold 8 of them. As we don't know which of these are faulty and faultless, we delivered them randomly. We expect the customer who takes a faulty machine will claim.
    (a) Give the probability mass function and the cumulative distribution function of the number of claims.
    (b) We need a spare part for each claim. How many spare parts do we need in order to satisfy all the claims surely.
    (c) And in order to satisfy all the claims with a probability of 0.8?

47. The number of visits until we get a new customer in a web page is a r.v. given by this function:

$$P[X = x] = 0.1 + \frac{k}{x} \; ; \; x = 1, 2, 3, 4$$

    (a) Calculate k, $P[X = x]$ to be a mass function.
    (b) Give mass function and cumulative distribution function as a table.

48. The distribution of the X random variable is given by:

$$P[X = x] = \frac{1}{k} \; ; \; x = 1, 2, \ldots, k$$

    Calculate k to be a mass function.

49. The distribution of the X random variable is given by:

$$F(x) = 1 - \left(\frac{1}{2}\right)^x \; ; \; x = 1, 2, \ldots$$

Give $P[X = 3]$, $P[X = 2]$, $P[X > 4]$ and $P[X < 6]$.

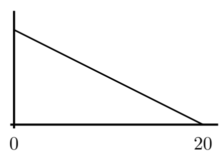
2.2 Continuous random variables

50. The percentage of students that pass an exam in given in this way:

$$f(x) = 2 - 2x \; ; \; 0 < x < 1$$

  (a) Draw the density function and interpret it.
  (b) What is the probability of being more than 10% of students passing the exam.
  (c) What is the probability of students passing the exam exceeding 50%.
  (d) Calvulate the probability of the percentage being in the %20-%30 interval. Warning: calculate for the closed and open intervals.
  (e) Give the distribution function and calculate by its means the previous probabilities.
  (f) Proof that given functions are true pdf and cdf.
  (g) If you want the 20% of students to pass the exam, which should be the pass note?

51. Sales in a store (thousands euros) follow this distribution:



  (a) Give the density function.
  (b) Give the cdf
  (c) Calculate the probability of the salens benig more than 10.000 euros, by means of both the pdf and the cdf.
  (d) Supposed independence between sales in different days, calculate the probability of sales exceeding everyday 10.000 euros over 5 consecutive days. Provided that really happened, which conclusion should be drawn?

52. A random variable follows this distribution:

$$f(x) = k - x \; ; \; 0 < x < k$$

  (a) Calculate k the previous function to be a pdf.
  (b) Give the cdf.
  (c) Calculate $P[0.5 < X < 1]$, by means of both the pdf and the cdf.

53. Daily production (kg) is rv, depending on the number of machines (k):

$$f(x) = \frac{2x}{100k^2} \; ; \; 0 < x < 10k$$

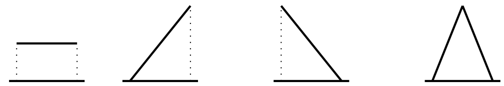
  (a) Proof that k is a parameter.
  (b) With 4 machines, what is the probability of production exceeding 30?
  (c) Give cdf, depending on k, and based on that, proof k is a parameter.
  (d) If product is packed by the kilogram, how many packs do we need, with 6 machines, in order to be the probability of packing all the production 0.8? And to be 0.9?
  (e) How many machines do we need in order to the probability of the production being more than 60 reaching 0.5?

54. The time to complete a task is given by the following cdf:

$$F(x) = \frac{x^3 - 8}{19} \; ; \;\; 2 \le x \le 3$$

  (a) Proof it's a cdf.
  (b) Give the pdf.
  (c) Calculate $P[X \ge 2.5]$ by means of both the pdf and the cdf.

55. Draw in a approximative way the cdfs corresponding to these pdfs:



56. The weight of an *clementina* orange follows this distribution:

$$f(x) = \frac{2}{10000}(x-100)\ ;\ 100 < x < 200$$

Calculate the probability of the weight being exactly 110 gr, both in a theoretical way and a practical way, known than balance gives 100-110-120-130-... weights. And 130 gr or less?

57. The number of customers entering a store follows this distribution:

$$f(x) = \frac{1}{100}\ ; \quad 200 < x < 300$$

What is the probability of entering exactly 250 customers? And 250-260 customers (closed interval)?

2.3 Expected value, variance and other moments

58. Sales in a month (units) distribute in this way:

Number	Probability
0	0.05
1	0.2
2	0.25
3	0.25
4	0.15
5	0.1
	1

Give the expected value and interpret it.

59. In a urn we have 20 faulty items, and 80 faultless. We extract 5 items randomly and at the same time. Give the expected value of the number of faulty items and propose a formula for calculating that value in general cases.

60. The number of items produced by a machine (thousands) distributes in this way:

$$f(x) = \frac{x}{2}\ ;\ 0 < x < 2$$

Plot the density function and approximate the expectation at first sight. Calculate it.

61. Monthly sales may be 1 and 2, with 0.4 and 0.6 probabilities respectively.
    (a) Which is the expected sales value for two months?
    (b) Variable cost and selling price per item are 100€and 200€, respectively. Which the expected gain value for one month, given that fixed costs are 50 €?
    (c) Through the last three months we sold 2, 1 and 2 items. Which is the average gain per month? Why is not the same as the expected value?

62. The tenperature in a freezer in any moment has the following distribution:

$$f(x) = \frac{1}{a}\ ;\ 0 < x < a$$

    (a) Plot the distribution and interpret it about the mean value.
    (b) Calculate the mean value and the standard deviation.
    (c) Calculate the 3rd central moment.

63. The number of faults in an item follows this distribution:

Number of faults	Probability
0	0.5
1	0.4
2	0.1
	1

(a) Calculate the 3rd raw moment.

(b) Calculate the expected value and the standard deviation.

(c) Calculate the 2nd central moment.

(d) Calculate the mean value and standard deviation of faults in 100 items, both in the case of independence and dependence betwen items about the numbers of faults.

64. The numbers of items produced in a factory may be 2 and 3 with 0.3 and 0.7 probabilities respectively. The next day, where the production the day before was 2, the production follows the same distribution; where the production was 3 yesterday, today may be 3 and 4 wtih 0.3 and 0.7 probabilities respectively. And so on the following days: whenever the maximum production happens, the next day we can produce that maximum and one more item; otherwise, the production is the same as the day before.

    (a) What is the probability of producing 12 or more items in the next 4 days?

    (b) Give the mean value of the production through the next 4 days.

    (c) If items are produced uniformly along the time, how much time do we need to produce 8 pieces on average?

65. There is just one worker in a factory. His production may be 1 or 2 pieces, with 0.4 and 0.6 probabilities. When the production is 1, the nex day we take another worker, who may produce 1 or 2 pieces, just like the first worker and with the same probabilities, but independently with him. When the production of the first worker in the first day is 2, we don't take the second worker and production follows the same distribution as the first day.

    (a) Give the expected production for those two days.

    (b) Give the expected time we need to produce 2 finished items.

2.4 Expected value and variance as criterions for decision

66. A firm has four investment choices. The gains of each investment are random and distribute in the following manner (negative gains are losses):

Gains	A investment	B investment	C investment	D investment
-2	0.05	0	0	0.05
-1	0.25	0.20	0.15	0.10
0	0.30	0.50	0.40	0.35
1	0.20	0.25	0.30	0.40
2	0.20	0.05	0.15	0.10

If needed, use this utility function:

$$U = \frac{\mu}{0.5\sigma} - 3P[loss]$$

    (a) Show that the utility function is correct.

    (b) Comparing two investments at once, discuss the preference for all the investments.

    (c) Sort the investments about the expected value, the risk and probability of loss and discuss which is the best investment.

67. We have to choose betwen two stock options, whose gains distribute in the following way:

    - A stock option:

$$f(x) = \frac{1}{2} - \frac{x}{8} \ ; \ 0 < x < 4$$

    - B stock option:

$$f(x) = \frac{1}{5}; \ 0 < x < 5$$

If needed, you should use this utility function:

$$U = \frac{\mu}{\sigma}$$

Discuss which is the best choice, in the long term as in the short term.

2.5 Chebysev's unequality

68. We don't know the exact distribution of the daily production, but we know that the mean is 100 items, and the standard deviation 20. What is the probability of the production being in the 60-140 interval?
69. The standard length for a given piece is 1000 mm. On average we comply the standard, with a 100 mm standard deviation. We accept the piece if the deviation from the standard value is less than 300mm. What is the probability of not accepting the piece?
70. Daily sales in a shop are 2000€ in a normal day, with a 200€ deviation. Whenever we predict sales will be more than 2600€, we will hire more workers. What is the probability of this happening?
71. The mean number of sandwiches requested in a day in a restaurant is 100. Deviation is 10. We need a bun for each sandwich. How many buns do we need in order of to be the probability of having enough buns to make all requested sandwiches 0.9? Idem, with a 0.99 probability. Idem, when the deviation is 20 (and the probability 0.99).
72. Daily sales in a shop are 1000€, on average, with a 200€ deviation. Bound the probability of being the sales 0f 3 days less than 2800.

3 Binomial distribution

73. The probability of a student passing an exam is 0.7.
    (a) In a group of 10 students, what is the probabilitaty of 6 students passing the exam?
    (b) And the probability of 6 students failing the exam?
    (c) And the probability of x students passing the exam?
    (d) Write in a simplified notation the distribution of the number of students passing the exam.
74. In a factory, we must produce 12 items for a customer. The probability of each item being faulty is 0.12.
    (a) Give the distribution of the faulty items among the 12 items.
    (b) Give the distribution of the faultless items among the 12 items.
    (c) What is the probability of having 4 faulty items?
    (d) What is the probability of having 3 faultless items?
    (e) What is the probability of having 3 faultless items and 9 faulty items?
    (f) What is the probability of having 2 faulty items or less?
    (g) What is the probability of having less than 3 faulty items?
    (h) What is the probability of having more than 8 faulty items?
    (i) Solve the last three questions by means of the Larson's nomogram.
    (j) What is the probability of having 10 faulty items or more?
    (k) What is the probability of the number of faulty events being between 4 and 6, both included?
    (l) Solve (c)-(k) questions with R software.
    (m) Which is the mean number of faulty items?
    (n) Solve by R: which is the most probable number of faulty items?
    (o) Solve by R: how many faulty items can we assure with a probability of at least 0.9 from the seller's point of view? (Solution)
75. 300 women eta 200 men live in a village.
    (a) We choose 15 personn randomly and with devolution. Give the distribution of the number of women.

(b) What is the probability of the number of women being just proportional to the number of women in the village?

(c) What is the meaning of the number of women in the latter question? How do you interpret the result?

(d) Give the probability of the number of women being 9 ± 2?

(e) Interpret the results in questions (b), (c) and (d).

(f) Give the probability of question (d), supposed you only know μ and σ and no that the number of women follows a binomial distribution.

76. In a given place the probability of raining is 0.4 and it's assumed total independence among different days. In order to construct a rooftop we need 7 days without rain. For how many days should we rent a crane to construct the rooftop with a 0.9 probability? And with 0.99 probability? (Solution)

77. We sell batches of 10 pieces. One of our customers inspects all the pieces and if he finds one faulty piece or more in one batch he rejects it. The probability of a piece being faulty is 0.15. On the other side, the customers buys 25 sets of batches along the year and if he rejects more than 20% of then, he will cancel the contract.

    (a) What is the probability of the contract being cancelled?

    (b) Which should the probability of a piece being faulty in order to be the probability of rejecting one batch at most 0.3? (Solution)

78. In a flight 25% of tickets are cancelled or become vacant. For a given flight we have 12 seats. How many tickets should we sell in order to be the probability of having overbooking at most 0.15?

3.1 Return periods in binomial distributions

79. The return period of having more than 200 mm rain in a day is 8 years. Which is more probable through 6 years: to have a rain of that magnitude of not to have it?

80. A bridge has been finished this year. The contractors think that a flood that would wreck the bridge should happen only once in 1000 years.

    (a) What is the probability of the bridge standing in the next 100 years?

    (b) For how many years will the bridge stand with a 0.95 probability?

3.2 Statistical testing related with binomial distribution

81. Normally 2% of pieces are faulty in a factory. Among the last 10 pieces 2 faulty pieces have been found.

    (a) Should we decide the production process is wrong? Significance level: 20%.

    (b) Test if the proportion of faulty items is exactly 10%. Significance level: 5%.

    (c) Test if the proportion of faulty items is exactly 30%. Significance level: 10%.

82. There are 20 questions in a test, each with 5 choices. A student has answered correctly 6 questions. Should we determine that he knew some of the questions or answered them randomly? Significance level: 10%.

83. 5 sellers (one of them is Peter) sold 15 machines in an exhibition. Peter sold 10 of them. Should we conclude that Peter is better than the other sellers? Significance level: 1%.

84. Last year there were 15 accidents in a road and 10 of them happened on Sundays and other holidays. Should we conclude accidents become more frequent on those days? Remark: there were 82 Sundays and other holidays last year. Significance level: 1%. (Solution)

85. There are 4 workers in a factory and each of them produces 6 pieces a day. Among the 24 pieces produced in a day we found 6 faulty pieces and only one of them was produced by Peter, the eldest worker. Should we decide he produces better that his colleagues? Significance level: 1%. (Solution)

86. A firm has designed a path to deliver packages. In order to program that in the most efficient way, the firm assumes that the probability for a package to be delivered at point A must be exactly 0.09. Among the last 14 packages, 4 of them were delivered at point A. Should we conclude that the program for the path should be remade? Significance level: 1%. (Solution)

3.3 Sign test

87. We collected some daily sales in a shop:

    46-87-56-64-55-67-64-65-72-75-98

    (a) Test if population median is 60. Significance level: 10%. Calculate also the p-value. (Solution)

    (b) Test if median is smaller than 70. Significance level: 10%.

    (c) Test if median is bigger than 80. Significance level: 10%.

88. We have collected data about the time needed by some workers to perform a given task, before and after a training course:

    Before: 46-87-56-64-55-67-64-65-72-75-98
    After: 48-67-62-60-53-60-65-60-68-72-81

    Has the training improved the time needed for the task? Significance level: 0.10.

89. We have collected these test scores among a group of students:

    67-78-75-81-82-77-72-69-87-75

    (a) Test if median is 80, by means of both the critical value and the p-value. $\alpha = 10\%$.

    (b) Test if median is bigger than 75. $\alpha = 10\%$.

    (c) After a training course, the same students got these scores:

    69-77-79-85-86-76-73-68-88-78

    Was the course profitable? $\alpha = 10\%$.

    (d) Set 90% and 80% confidence intervals for the median. (Solution)

3.4 Geometric distribution. Negative binomial distribution.

90. The probability of a faulty piece being produced is 0.12.

    (a) Which is the distribution of the number of produced faultless pieces before the first produced faulty piece?

    (b) What is the probability of having 6 faultless pieces just before the first faulty piece?

    (c) What is the probability of having 2 faultless pieces or less just before the first faulty piece?

    (d) What is the expected number of faultless pieces before the first faulty piece?

    (e) What is the probability of having 2 faulty pieces before the first faultless piece?

    (f) What is the expected number of faulty pieces before the first faultless piece? (Solution)

91. The probability of a faulty piece being produced is 0.12.

    (a) Which is the distribution of the number of produced faultless pieces before the third produced faulty piece?

    (b) What is the probability of having 6 faultless pieces before the third faulty piece?

    (c) What is the probability of having 2 faultless pieces or less just before the third faulty piece?

    (d) What is the expected number of faultless pieces before the third faulty piece?

    (e) What is the probability of having 2 faulty pieces before the fourth faultless piece?

    (f) What is the expected number of faulty pieces before the fourth faultless piece?

4 Hypergeometric distribution and Fisher's exact test

92. In an urn we have 20 faulty and 80 faultless pieces. We take 5 pieces randomly at the same time.

    (a) Calculate the probability of having 3 faulty pieces among those 5 pieces, both by means of the binomial coefficientes and multiplying simple probabilities.

    (b) Calculate the mean number of faulty pieces.

(c) Calculate the probabilities of all possible values for the number of faulty pieces. Calculate the mode by both examining the previous results and applying the formula for the mode.

(d) State the probabilities equal to the probability of having 3 faulty pieces (that is to say, the symmetries)

(e) Calculate the variance of the number of faulty pieces, and compare it to the variance of faulty pieces, when pieces are drawn with devolution.

(f) Calculate the probability of the faulty pieces being 60%, both when we draw 5 pieces and 25 pieces. Do we have a paradox with these results? (Solution)

93. *Lady tasting tea experiment*: A lady usually drinks tea with milk. We want to know if she is able to distinguish if milk was poured before or after the tea. She is blinded and we give her 8 tea cups, 4 with the milk poured before and the other 4 after. The lady guessed twice that milk was poured before, and 3 times that was poured after. Using Fisher's exact test, analyze if lady is able to guess if milk was poured before and after. $\alpha = 0.1$.

94. We have asked some people about sex and political attitude:

Sex ↓ / Attitude →	Conservative	Leftist	Total
Man	6	8	14
Woman	11	4	15
Total	17	12	29

(a) Use Fisher's exact test to analyze if women are more frequently conservative than men. $\alpha = 0.1$.

(b) Perform the test changing the pivot frequency.

5 Poisson distribution

95. In a machine 4.2 failures happen on average per hour and randomly.

(a) What is the probability of having no failures in 2 hours?

(b) Calculate the probability of having respectively 2, 3, 4 eta 28 failures in one hour. Give the interpretation of the results.

(c) We must perform a task in 22 minutes. What is the probability of completing the task without any failure?

(d) What is the probability of having 3 failures or less in 2 hours?

(e) What is the probability of having 8 failures or more in 2 hours?

(f) What is the probability of having 10 to 15 failures in 2 hours, both included?

(g) What is the probability of the time between 2 failures being at least 2 hours?

(h) How many failures will be in an hour with a 0.9 probability? And with a 0.99 probability? Give the answer from both the seller's and the production manager's point of view.

96. Hepatitis-C infected patients admitted in a hospital are 8.6 a week on average, randomly and independently one from another. We need one dose per patient to treat the illness.

(a) How many doses must we prepare in order to treat all the patients with a 0.9 probability?

(b) Idem, with a 0.99 probability.

(c) In two weeks 24 patients were treated. Should we conclude that the prevalence of the illness has increased? Significance level: 0.01.

(d) As health managers are worried about the growth of the number of patients, thay have prepared a prevention campaign. Following the campaign, along the next 4 weeks, there were 14 patients. Should we state that campaign was successful? Significance level: 0.01.

97. On average 6 customers per minute enter a shop, randomly and with independence from one to another. Each customer needs 2 minutes to pay. How many cashiers should we hire in order to have a 0.9 probability of not having a queue? (Solution)

5.1 Poisson distribution as limit of the binomial distribution

98. We are managing industrial process, the probability of producing a faulty item is 0.0022. We have sold a batch of 4000 items, but we'll return it whenever we find more than 6 faulty items.
    (a) What is the probability of returning the batch? Calculate both by means of the binomial distribution and the Poisson distribution. What is the meaning of lambda in this case?
    (b) Among 1000 items, how many faulty items can we ensure with a 0.9 probability from the seller's point of view?
    (c) How many faulty items can be assured with a 0.9 probability from the customer's point of view ? (Solution)
99. The probability of a worker having an accident in a year is 0.0012.
    (a) On average, how many accidents happen among 1000 workers?
    (b) In a factory there are 3000 workers and there were 7 accidents last year. Should we conclude that prevention is not as successful as in the other factories? Significance level: 0.05. (Solution)

5.2 Return periods in Poisson distributions

100. The return period of having more than 200 mm rain in a day is 8 years.
    (a) Which is more probable through 6 years: to have a rain of that magnitude of not to have it?
    (b) How many years are needed to balance the probabilities of having heavy rain and not having it?
101. A bridge has been finished this year. The contractors think that a flood that would wreck the bridge should happen only once in 1000 years.
    (a) What is the probability of the bridge standing in the next 100 years?
    (b) For how many years will the bridge stand with a 0.95 probability? (Solution)

6 Exponential distribution

102. A machine stops twice an hour, randomly and independently.
    (a) What is the probability of the time between consecutive stops being larger than 20 minutes? Calculate both in terms of minutes and hours.
    (b) What is the probability of the time to the next stop being larger than 20 minutes?
    (c) What is the probability of that time being 20 to 30 minutes?
103. The mean time between failures in a machine is 10 minutes, and those failures happen randomly and independently.
    (a) What is the probability of the time between consecutive stops being larger than 5 minutes?
    (b) Given that it is 60 minutes that the last stop happened, what is the probability of the time to the next stop being larger than 5 minutes? Express the probability formally and interpret the previous result.
104. On average 4 customers visit a website in an hour. The website shuts down for 40 minutes. What is the probability of losing a customer in that interval? Calculate it both by means of the number of customers and the time between customers.
105. Customers come randomly and independently to a queue, with a mean time till the next customer of 5 minutes. It turns out that the next customer has come in 20 minutes. Should we reject the 5 minutes average? Significance level: %1.

6.1 Gamma distribution

106. The number of arrivals to a service occur randomly, at a 1.4 per minute rate.
    (a) Express the distribution of time till the 4th arrival and its mean and variance.
    (b) What is the probability of passing more than 15 minutes till the 4th arrival?
107. Time between customers occurs randomly, with an average time of 4 minutes.
    (a) What is the probability of passing more than 5 minutes till the 2nd customer?

(b) What is the probability of passing less than 15 minutes till the 6th customer?

(c) Give the mean and variance of the time till the 6th arrival.

7 Uniform distribution

7.1 Discrete uniform distribution

108. The number of daily sales of washing machines at a retailer's is between 0 and 9, following a discrete uniform distribution. Calculate:

    (a) The probability of selling on a given day 3 machines or less.

    (b) The mean number and variance of machines sold.

    (c) The probability of a given sequence of sales along 4 days (e.g., 2,6,1,8). Give the number of days, the minimum and the maximum for that sequence.

    (d) Along 4, 5 and 6 days respectively, the probability of the maximum being exactly 8, as in the given sequence.

    (e) The shop assistant is able to manage only 7 machines one day. Give the probability of not being able to manage all the sales at least one day along the 4 days.

    (f) Along 4, 5 and 6 days respectively, the probability of the minimum being 1, as in the given sequence.

    (g) There are losses when the number of daily sales is smaller than (or equal to) 2. Give the probability of having at least one day along those 4 days of being losses.

    (h) How many days do we need in order for the probability of the maximum being 9 to be at least 0.95? And in order for the probability of the maximum being 9 or less to be at least 0.95?

    (i) How many days do we need in order for the probability of the maximum being 8 to be at least 0.95? And in order for the probability of the maximum being 7 with the same probability?

    (j) Which should be the top value of the distribution, instead of 9, in order for the maximum along 4 days to be at least 15 with a 0.95 probability? Idem, with a 0.8 probability. And along 8 days with both probabilities? (Solution)

109. We have composed a list fron no.1 to no.222 with all the students of our college. We take a random simple sample. What is the probability of drawing student no.26, 36, 67 and 197. Do it with devolution and without devolution.

110. An artist's engravings are numbered, but the size of the engraving set is unknown. We know there are engravings no. 10, 34, 23 and 8.

    (a) Give an estimation about the engraving set size.

    (b) If the 34 maximum had happened with 10 engravings, instead of 4, which should the estimation for the set size? Why do you think it's smaller than the previous result?

7.2 Continuous uniform distribution

111. We only know the price growth of an item will be 0 to 10 in the next year.

    (a) What is the probability of the price growth being bigger than 6 (give the answer both by reasoning and calculation). Plot the that probability into the density function.

    (b) Which is the expected price growth? Give the answer both by reasoning and calculation.

    (c) Which is the variance of the price growth? (Solution)

112. $X \sim U(0, 8)$. Calculate z, $P[X > \mu_X + z] = 0.1$. (Solution)

113. Proof that the conditional distribution below a given value for a uniform distribution follows also a uniform distribution.

114. Sales in a shop distribute between 100 and 200, and furthermore we have absolute uncertainty about the sales number in that interval.

    (a) Simulate sales numbers along 4 days.

    (b) Give the maximum, minimum and range for each 4 days sequence simulated previously.

    (c) Repeat (a) and (b) 20 times.

(d) Depict maximum, minimun and range values in a suitable chart.

(e) Interpret the charts: which are the mean values, appproximately, for maximum, minimum and range along 4 days?

(f) Calculate exact expected values for maximum, minimum and range along 4 days.

(g) Explain the previous expected values in a intuitive way.

(h) Predict and calculate the probability of the maximum being smaller than 140.

(i) Predict and calculate the probability of the minimum being larger than 180.

(j) Predict and calculate the probability of the range being smaller than 10. (Solution)

115. Sales in a shop distribute $U(0, b)$, b being an unknown parameter. To estimate b, we take the maximum of sales along n days. E.g., if we get 23-86-176-158, we estimate $\hat{b} = 176$. It's clear that the estimation, that is to say, the maximum will always be below b, so we have a systematic error. But we can control it.

(a) In percentage, how much error do we make taking the maximum sales number along 7 days as an estimation for b?

(b) Along how many days should we collect sales data if we want the error to be at most 5%?

(c) What should we do to fix that error?

(d) We assume that sales distribute $U(100, 200)$. Along 4 days the maximum was 140. Should we conclude that the maximum of the distribution is really 200? $\alpha = 0.1$ (Solution)

8 Power law distributions

116. We have collected data about family incomes in a village:

Income	Families
100-200	86
200-400	44
400-600	20
600-1000	14
1000-2000	6
2000-5000	2

(a) Plot incomes and above probability, and decide if power law distribution is a suitable model.

(b) Make an estimate for α parameter, taking these exact values, drawn randomly among the families: 122-134-148-168-186-254-386-875.

(c) Calculate theoretical probabilities with the previous estimate.

(d) Matching the expected value with the arithmetic mean calculated with previous data, make a new estimate for α parameter.

(e) Which is the best fitting estimate?

(f) Taking the best estimate, calculate the percentage of wealth owned by the richest 10% of families.

117. We have collected data about the best seller books last month:

Book	Sales
A	86
B	52
C	32
D	22
E	16

(a) Examine if Zipf'law is suitable for those data,

(b) Give two estimates for the parameter of the Zipf's law and examine which is the best fitting one.

(c) Shortly we will receive the absolutely best seller book in the last years. If the sales of the other books remains constant (if Zipf' law remains constant), how many books will we sell from that new book?

9 Normal distribution

118. Tenperature into a freezer follows a $N(0,1)$ standard normal distribution. Calculate these probabilities both by means of the statistical table and R programming language commands:

    (a) $P[Z < 1.42]$

    (b) $P[Z < 3.98]$

    (c) $P[Z < 5.6]$

    (d) $P[Z > 2.75]$

    (e) $P[Z < -0.68]$

    (f) $P[Z > -1.02]$

    (g) $P[0.48 < Z < 1.92]$

    (h) $P[-1.24 < Z < -0.98]$

    (i) $P[-2.19 < Z < 0.55]$

    (j) Give the value in the standard normal distribution that accumulates below it a 95% probability.

    (k) Give the value in the standard normal distribution that accumulates beyond it a 20% probability.

    (l) Give the value in the standard normal distribution that accumulates below it a 10% probability.

119. Daily production (P) in a factory follows a $N(68kg, 4kg)$ distribution. Calculate these probabilities:

    (a) $P[P < 77]$

    (b) $P[P > 62]$

    (c) $P[64 < P < 72]$

    (d) Calculate the maximum production in 99% of the days.

    (e) How much production can we ensure with a 0.9 probability?

    (f) Known that production in one day has been 54 kg, should we conclude that production has dropped?

    (g) Discuss how to calculate this probability in practice

120. Grade in a test follows the normal distribution, with an average of 110 poins and a deviation of 5 points. In order to pass the test, the students must achieve 120 points.

    (a) What is the probability of passing the test?

    (b) We exclude students below 95 points. How many students do we exclude as a percentage?

    (c) Calculate the number of points to exclude 30% of the students.

    (d) Give a 90% symmetric confidence interval for the number of points.

121. Daily production in a factory follows the normal distribution, with a 8 ton average and a deviation of 1 ton.

    (a) What is the probability of producing less than 30 tons in 4 days?

    (b) How much production can we ensure with a 90% probability?

    (c) How many days do we need to fulfill a batch of 60 tons, if we want the probability of not fulfilling the deadline to be 15%?

    (d) Idem, with a 1% probability.

    (e) Taxes for producing will be 40.000€, with a tax allowance of 1.000€per ton. What is the probability of paying for taxes more than 10.000€? (Solution)

122. 12 people get into an elevator, and the maximum weight allowed is 900 kg. Which should the average weight of a person in order to not exceeding it being 0.9? Remark: standard deviation of the weight of a person is 10 kg. (Solution)

123. *(Overview problem)* It's assumed that daily sales in a shop distribute according to the gaussian law, with an average of 2000€ and a standard deviation of 400€ .

 (a) Plot the distribution and draw there the probability of selling more than 1400€ .

 (b) Calculate the previous probability.

 (c) Calculate the minimum amount of the sales on 99% of the days. Idem with a 90% percentage.

 (d) Calculate the probability of selling less 8200€ along 4 days.

 (e) How many days do we need to sell more than 30.000€with a 95% probability?

 (f) Daily taxes are a fixed amount of 200€ plus a 10% over the sales. What is the probability of the taxes being bigger than 420€ ?

9.1 De Moivre-Laplace theorem

124. In a production process the probability of a faulty item is 0.25. We have a 100 item batch.

 (a) What is the probability of having 30 faulty items or less?

 (b) How many faulty items are expected to be?

 (c) What is the probability of having exactly the number of faulty items expected? Interpret the result.

 (d) How many faulty items can we ensure with a probability of 90%? (Solution)

125. 1000 people are attending a recruitment process. The probability of passing the first exam is 0.6. How many chairs should we get if we want the probability of all the candidates in the second exam having a chair available to be at least 0.99? (Solution)

126. Daily production follows a N(100kg, 10kg) distribution. What is the probability of having in one year (365 days) at least 317 days with a production of at most 110kg? (Solution)

9.2 Normal approximation of the Poisson distribution

127. The expected daily number of failures in a machine is 0.16, following the Poisson distribution. We need one piece to fix every failure.

 (a) What is the probability of having more than 50 failures in a year?

 (b) How many pieces do we need in order to fix all the failures in a year with a 0.99 probability?

 (c) With 100 pieces, for how many days can we ensure that we have enough pieces to fix all the failures with a 0.95 probability? (Solution)

128. Following a Poisson process, 1.4 batteries are exhausted on average each day in a machine.

 (a) Throughout 40 days, what is the probability of exhausting more than 50 batteries?

 (b) How many batteries should we get in order to have enough energy for 80 days with a 0.9 probability? (Solution)

9.3 Central limit theorem

129. Substance consumed in a day follows an uniform distribution, in a 5 to 10 liters interval.

 (a) What is the probability of consuming less than 420 liters in 40 days?

 (b) How many liters must we get for 40 days in order to being the probability of having enough substance 0.99?

 (c) Idem, if the consumption follows a U(0,15) distribution.

 (d) Provided we have 500 liters, for how many days do we have enough substance with a 0.99 probability? (Solution)

130. The share price possible daily increments are +1, 0 and -1€ with respectively 0.2, 0.5 and 0.3 probabilities.

 (a) After 100 days, what is the probability of not losing?

 (b) How much money should we have after 100 days in order to be able to pay the losses with a 0.99 probability? (Solution)

131. 1.4 batteries are exhausted each day on average following the Poisson distribution.

(a) Provided we have 40 batteries, for how many days do we have enough energy with a 0.99 probability?

(b) How many batteries must we get provided we want the probability of having enough energy for 80 days being 0.98?

(c) Which should be the average duration for each battery, if we want to reach a total duration of 60 days with 40 batteries with a 0.9 probability?

132. In a factory the daily average production from Monday to Thursday is 146 tons, with a standard deviation of 10 tons, following an unknown distribution. On Fridays and Saturdays, the factory works only along the morning, and so the average production is 64 tons with a standard deviation of 6 tons.

(a) If along the next 8 weeks we are working only from Monday to Thursday, how much production can we guarantee for those days with a 0.96 probability?

(b) Calculate the probability of less than 7000 tons along 10 weeks.

(c) Give the deadline or number of weeks needed to produce 8000 tons with a 0.96 probability? (Solution)

133. The number of failures in a machine may be 0, 1, 2 with 0.2, 0.5, 0.3 probabilities respectively. Following a maintenance plan, in 100 machines there have been 90 failures. May we conclude that the maintenance plan has been successful? Significance level: 2%.

134. In a given flight, we assume that the weight of each baggage is on average 20kg, with a 3kg deviation. The exact distribution is unknown.

(a) If the flight gets 100 travelers, what is the probability of having more than 2100kg baggages?

(b) Which amount of baggages can you ensure with 96% probability?

(c) How many travelers can we take with the condition that not more of 2000kg will be taken with a 0.99 probability?

(d) Calculate the probability of not having more than 100kg with 5 baggages. Why can't you apply CLT in ths case? What should you apply? And what if weight of baggages were normal distributed?

(e) What is the probability of the sample mean of 50 baggages being bigger than 21?

(f) If the sample mean of 50 baggages was really bigger than 21kg, which should your decision about the theoretical average weight? Significance level: 0.05 (Solution)

135. Daily production in a factory follows a U(10,20) distribution, in kilograms. In 50 days, the average production has been 14 kg.

(a) Should we conclude that production has decreased? Significance level: 2%.

(b) Which is the production level in order to claim that average production has decreased on average? Significance level: 2%.

(c) Calculate the critical values in order to reject the null hypothesis with sample sizes of 100 and 500 days. Interpret the results.

(d) Which is the production level in order to claim that average production has just changed? Significance level: 2%. (Solution)

136. A product is the increasing phase of its life-cycle. We think that sales per months distribute in this way $U(800 + 200i, 1800 + 200i)$ (i: rank of the month from nowadays).

(a) Calculate the average sales level and the probability of selling more than 1600€ on each of the next three months. Interpret the results in the light of the statement of the problem.

(b) Give the probability of selling more than 180.000€ along the next three years.

10 Statistical inference: validation

10.1 Goodness of fit: chi-square test

137. 60 customers tasted 4 recent marketed yoghourts and were asked about their favourite one. A, B, C and D labeled yoghourts were chosen 20, 14, 12 and 14 times respectively. Should we conclude that yoghourts are equally probable to be chosen? Significance level: 10%.

138. We collected throughout 100 days the daily number of failures in a machine:

Number of failures	0	1	2	3	>3
Number of days	21	19	15	20	25

Can we conclude that the number of failures follows a Poisson distribution? Significance level: 10%.

139. Scores obtained by a group of students are given below (original data: 21, 46, ...):

Score	0-20	20-40	40-60	60-80	80-100
Number of students	2	14	34	38	12

Test whether data follow a normal distribution. Significance level: 10%.

140. A restaurant serves menus from Monday to Friday, at noon and evenings. The total number of menus served is given below:

Menus	Monday	Tuesday	Wednesday	Thursday	Friday	Total
Midday	38	45	38	58	40	219
Evening	18	31	26	30	36	141
Total	56	76	64	88	76	360

(a) Conclude whether number of menus served from Monday to Friday follows an uniform distribution, by means of the chi-square test. Significance level: 5%.

(b) Test whether menus served at noon are twice menus served at evening. Significance level: 5%.

(Solution)

141. Times in days till a failure in a machine are given below:

$26.2, 22.3, 33.5, 19.0, 24.7, 25.6, 26.2, 28.9, 27.6, 26.5, 27.1$

$32.4, 36.2, 34.1, 28.7, 26.5, 25.4, 23.4, 21.6, 22.0, 20.6, 30.2$

Setting < 19, $19-21$, $21-23$, ... intervals, test whether data follow a normal distribution, by means of the chi-square test, provided that we must previously estimate the mean and the standard deviation. Significance level: 10%. (Solution)

142. Times in days till a failure in a machine are given below:

$26.2, 22.3, 33.5, 19.0, 24.7, 25.6, 26.2, 28.9, 27.6, 26.5, 27.1$

$32.4, 36.2, 34.1, 28.7, 26.5, 25.4, 23.4, 21.6, 22.0, 20.6, 30.2$

Setting $0-10$, $10-20$, $20-30$, $30-40$ eta >40 intervals, test whether data follow an exponential distribution, by means of the chi-square test, provided that we must previously estimate the mean of the distribution. Significance level: 10%. (Solution)

10.2 Testing randomness and independence: the runs test

143. Allegedly we have collected some random data about the production in a factory. The data are given below:

$34.6 - 26.2 - 31.0 - 28.7 - 29.5 - 33.4 - 35.6 - 27.2 - 30.8 - 28.9$

$36.1 - 27.5 - 34.5 - 29.7 - 35.6 - 32.8 - 28.8 - 32.3 - 27.2$

Apply the runs test to conclude if data were collected independently. Significance level: 5%.

144. We have compiled the share price's change in the last days:

$$\downarrow\uparrow\uparrow\downarrow\downarrow\downarrow\uparrow\downarrow\downarrow\uparrow\uparrow\downarrow\uparrow\uparrow$$

Should we conclude they happen randomly and independently? Significance level: 5%.

145. We have compiled producted pieces have been faulty or faultless in a production process:

$$OOOOOXXOXXXOOOOXOOOOOOOOXXXOOXOXOOOOOOOXOOOOXXX$$

Test whether the process develops independently and drawn pieces provide a random sample by means of the runs test. Significance level: 10%.

10.3 Testing homogeneity: Wilcoxon rank-sum test

146. We have compiled califications given by two teachers in a exam:

A teacher	B teacher
5.4	6.5
6.2	7.4
8.7	4.7
9.5	5.6
7.6	5.4
8.2	8.6
3.5	7.2
7.8	

Should we conclude that both teachers give the califications in the same way? Significance level: 5%. Hint=1+2+...+15=120.

147. We have compiled some invoice amounts paid in a shop by sex (15 mean and 20 women):

$Men : 3, 3, 5, 6, 8, 10, 10, 11, 11, 12, 12, 12, 16, 19, 20$

$Women : 2, 7, 9, 11, 13, 13, 15, 17, 17, 18, 20, 21, 23, 24, 25, 25, 27, 32, 36, 39$

Test with those data whether men and women have the same buying behavior. Significance level: 5%. Solve it looking into the corresponding tables as well as by the normal approximation. Hint: 1+2+...+35=630. (Solution)

148. Children with math comprehension problems were given a special training program last year. They carried a test before and after the program. The results are given below:

$Before : 22, 32, 43, 28, 27, 36$

$After : 25, 42, 50, 35, 35, 42$

Test whether the program has been successful. Significance level: 5%. Hint: 1+2+...+12=78. Remark: surveyed children before and after the program are different , so we have independent samples. If the children were the same, we would have dependent samples (paired samples) and hence we would have to perform another kind of test (sign test, for example). (Solution)

149. Data about number of movie tickets sold on Saturdays (15) and Sundays (12) are given below:

Saturday : $126 - 91 - 68 - 122 - 113 - 137 - 111 - 86 - 100 - 82 - 96 - 121 - 97 - 95 - 89$

Sunday : $81 - 98 - 129 - 101 - 121 - 124 - 133 - 108 - 84 - 89 - 86 - 131$

Test if more tickets are sold on Sunday, by means of Wilcoxon rank-sum test and using the normal approximation (without tables). Significance level: 1%. Hint: 1+2+...+27=378. (Solution)

11 Sampling distributions

150. A population is composed by 1-2-3 elements. We draw a sample of size 3, with devolution. Give the sampling distribution of the arithmetic mean and interpret it.

12 Parametric tests

12.1 Tests for the mean

151. Sales in a shop follow a normal distribution. We suppose the standard deviation is 100. In a normal day sales are 1000 on average. Throughout 10 days, average sales have been 1100. Should we conclude that averages sales have increased? Solve both by the p-value and the critical region. Significance level: 5%.

152. We have drawn these data from a normal population: 22-26-24-24-25-23.

   (a) Test whether the population mean is 22. Significance level: 10%.

   (b) Without any other calculation, which other hypothesis should we accept or reject from the previous results?

   (c) Given 8 data, sample mean has been 28 and sample variance 2. Should we conclude that population mean has increased? Significance level: 1%.

153. It has been set that a null hypothesis establishes that a given flight is 300 minutes on average. We know that standard deviation is 40 minutes, baina we cannot establish the model for the data. Provided that sample mean is 310 minutes for 100 flights, test the given hypothesis by both the p-value and the critical region. Significance level: 5%.

154. We have drawn 50 data about time following an alleged exponential distribution. Normally mean time is 100 minutes. Set the test in order to conclude whether mean time has decreased. Significance level: 5%. (Solution)

155. We have compiled water comsumption data from 80 families: $\overline{x} = 7$; $\sum x_i^2 = 4800$. The model for data is supposed to be a gamma distribution, with unknown mean and variance.

    (a) Decide by means of the p-value if population mean may be bigger than 7.5. Significance level: 1%.

    (b) Now decide if population mean may be smaller than 7.5. Significance level: 1%.

    (Solution)

156. A component's duration follows a normal distribution, with a standard deviation of 4 days. Quality specifications require mean duration to be at least 10 days. Sample mean from 9 data has been 8 days. Test whether the specification is met or not. Significance level: 1%. (Solution)

157. A machine's output is 40 units on average, and allegedly follows a normal distribution. In order to have independent data, we have drawn 4 output data from different days: 38-39-35-36. What should we decide with those data about the average output? Significance level: 5%.

158. (a) The quality standard established for the resistence of iron produced in a given factory is $100kg/cm^2$. We have taken iron samples from 8 independent produced batches and got these results:

$$99 - 98 - 96 - 102 - 101 - 100 - 95 - 96$$

    We suspect that the standard is not met on average. Test the data in order to take a decision about that.
    (b) On the other side, we have another iron feature: flexibility. Quality standards establish that flexibility must be exactly 200 units on average (too rigid iron breaks easily and too flexible is not useful), with the standar deviation for the population being 10 units. Give the decision rule to test whether the specification is met on average, that is to say, test $H_0 : \mu = 200$.

12.2 Tests for the proportion

159. In a market, 26% of the consumers buys our product. This year 44 consumers of 200 have declared they will follow to buy our product. With those data, draw a conclusion about the change in the proportion of consumers, by means of both the p-value and the critical region. Significance level: 5%.

160. Our supplier claims that at most 4% of pieces are faulty. We suspect that it's not true. In order to test our suspicions, we draw randomly and inspect 200 pieces. With a 10% significance level, how many faulty items must we have in order to reject the supplier's claim? (Namely, calculate the critical value) Idem, with 5% and 1% significance levels. Build a table with the results and draw a conclusion.

12.3 Tests for the population variance.

161. We have drawn 10 data from an alleged normal distribution. Sample variance (without correction) is 36. Test whether the standard deviation in the population is 5 or less. Significance level: 0.01.

162. We have compiled the number of microorganisms in two pills, in hundred of millions: 2.2, 4.2. Quality specifications require that the variance of that number must be 2 or less.

    (a) Why are small variances required in order to have a good quality?

    (b) Should we decide that the given specification is fulfilled? Significance level: 0.10.

13 Confidence intervals

13.1 Confidence intervals about the mean

163. We have compiled the duration of a bus tour (in minutes)

$$12 - 11 - 13 - 16 - 18 - 20 - 10 - 14$$

    Duration follows an alleged normal distribution, that should be validated by a goodness-of-fit test.

(a) Estimate the mean and the standard deviation. Are they good estimates?

(b) Give a 90% confidence interval about the population mean.

(c) Idem, with a 99% confidence level.

164. We have compiled sales number in a shop throughout 10 Mondays. Sample mean is 282 currency units and corrected standard deviation 32. A normal population is assumed.

(a) Why do we take only sales on Mondays?

(b) Give a 95% confidence interval about the population mean.

(c) If we had drawn 20 data, instead of 10, which would be the confidence interval? Interpret the new result.

(d) If standard deviation had been 44 with 10 data, which would be the new interval? Interpret the result.

165. We have compiled some outputs per hour in a machine:

$$56 - 44 - 48 - 52 - 60$$

(a) Assumed a normal distribution, give the 90% confidence interval about the population mean.

(b) Which 90% confidence interval should give the machine dealer?

(c) Which 95% confidence interval should give the production manager?

166. We a drawn a random sample of 86 data from a normal population. Sample mean is 144 units, and sample variance 121 units. Give the symmetrical 98% confidence interval.

167. We have collected daily outputs during 10 days, with these results:

$$\sum_{i=1}^{10} x_i = 108 \; ; \sum_{i=1}^{10} x_i^2 = 1234$$

(a) Give a point estimation and a 99% symmetrical confidence interval about the population mean.

(b) May we conclude that the average output is bigger than 10 units with a 90% confidence?

13.2 Confidence intervals about the proportion

168. 120 faulty items have been found in 1000 items.

(a) Give 90% ad 99% confidence intervals about the proportion.

(b) Which would be the confidence interval if 1200 faulty items had been found in 10.000 pieces? Interpret the new result in connection with the latter.

169. 16 faulty items have been found in 256 items.

(a) Give the error in the estimate of the proportion with a 80% confidence.

(b) Which interval should give the machine dealer with the same confidence?

(c) Which interval should give the customer in order to make a complaint with the same confidence?

170. We have compiled the amount of a given substance in some batchs of a product (mg/l):

22.0 – 17.7 – 23.0 – 22.7 – 21.5 – 20.6 – 16.2 – 28.7 – 27.4 – 15.6 – 28.4 – 20.1 – 17.1 – 19.5 – 18.6
24.0 – 22.6 – 22.9 – 24.5 – 25.4 – 24.4 – 20.7 – 14.6 – 18.3 – 23.6 – 20.5 – 24.8 – 22.9 – 23.4 – 25.2
24.6 – 24.4 – 26.0 – 26.8 – 21.2 – 13.4 – 20.2 – 17.4 – 25.4 – 23.6

When the amount of substance is bigger than 25.5 mg/l, we reject the batch.

(a) Give an estimate of the proportion of rejected batches and calculate the corresponding error with a 96% confidence.

(b) Give the corresponding symmetrical confidence interval.

(c) We want to claim the product has a good quality. Which confidence interval should we give with the same confidence level?

171. In a election, we want to estimate the proportion of the voters for X party with a 90% confidence, with an error of ±2%.

    (a) How many voters should we poll or survey for that? Be cautious or pessimistic.

    (b) Give the new sampling size if the error switches to ±1%? Be cautious.

    (c) With a ±1% error, if the number of voters for X party are eventually 2.000, give the corresponding confidence interval.

172. Provided that in a pilot sample we got 22 faulty items in 100 pieces.

    (a) Give the sampling size for a 99% interval confidence interval for a proportion of faulty items with a 4% error

    (b) Idem, with a 90% interval.

    (c) Idem, being cautious about the sample size.

    (d) Why is the sample size in the previous section bigger?

Solutions for selected problems

Gizapedia

gizapedia.hirusta.io

Handout on Combinatorics

1 Is it a choosing problem or an ordering problem?

2 If it's a choosing problem ...

(a) Give some examples about the objects or instances to count, using a proper coding (letters or/and numbers).

(b) Give n: the number of elements (letters or numbers) we have to create the object.

(c) Give k: the size of the object.

2.1 Using the proper formula in choosing problems

2.1.1 Is order relevant?

Give different examples with the same elements in different ordering and ask yourself: are there the same? For example, are ab and ba different?

- Yes: order is relevant: order YES
- No: order is not relevant: order NO

2.1.2 Is it possible to repeat the elements in the object?

Give some examples with repeated elements and ask yourself: is it possible? For example, is aa possible?

- Yes: repetition YES
- No: repetition NO

2.1.3 Applying the proper formula

FORMULAE	Repetition no	Repetition yes
Order yes	$V_n^k = \frac{n!}{(n-k)!}$	$VR_n^k = n^k$
Order no	$C_n^k = \binom{n}{k}$	$M_n^k = \binom{n+k-1}{k}$

3 If it's a ordering problem ...

(a) Give some examples about the objects or instances to order, using a proper coding (letters or numbers).

(b) Are always the elements (letters or numbers) all **different?**

(c) If they are, use **permutation** formula: $P_n = n!$

(d) If they are not, use **permutation with repetition** formula: $P_n^{\alpha,\beta,\dots} = \frac{n!}{\alpha!\beta!\dots}$

1 Combinatorics

7. **You have to choose a commission of 2 students between 10 students. Order is not relevant. Two students refuse to be together in the commission. How many ways are there to choose the commission? Idem if the commission comprises 4 students.**

It's a choosing problem *with a restriction*. Let's suppose that the persons who don't want to be together are a and b.

First, we don't take into account the restriction. Into the commision the order is not relevant and there must be two different persons. Hence:

- Order: No.
- Repetition: No.

So, we must use the combinations formula: $C_n^k = \binom{n}{k} = \dfrac{n!}{k!(n-k)!}$

In order to apply the formula:

- we can choose among 10 students. So, $n = 10$.
- we must give couples. So, $k = 2$.

Hence: $C_{10}^2 = \binom{10}{2} = \dfrac{10 \times 9}{2!} = 45$

But in this manner we are including commissions with a and b together. So, from 45 we must substract those commisions with a and b together. How many are there? Just one: ab.

So, the answer is 45-1=44.

If the commission has 4 students, the number of commisions without any rectriction is:

$C_{10}^4 = \binom{10}{4} = \dfrac{10 \times 9 \times 8 \times 7}{4!} = 210$

But among those 210 commisions we have some with a and b together. How many?

If a and b are together, we can choose any 2 other students among the other 8 students:

$C_8^2 = \binom{8}{2} = \dfrac{8 \times 7}{2!} = 28$

So the answer is 210-28=182.

We can solve the problem in another way, adding ways to create the commission.

Removing those instances with a and b together, we can find two types of commissions: those without a and b, and those with a or b:

- Commissions without a or b: we must choose 4 people among 8 people: $C_8^4 = \binom{8}{4} = \dfrac{8 \times 7 \times 6 \times 5}{4!} = 70$
- Commissions with either a or b: we must choose 3 people among 8 people $C_8^3 = \binom{8}{3} = \dfrac{8 \times 7 \times 6}{3!} = 56$, but after that we must choose to include a or b, 2 options, so the total number of commissions in this manner is $56 \times 2 = 112$

Hence, the total number of commisions is $70 + 112 = 182$, the same result as the latter.

2 Laplace's rule

10. **A postman has to deliver 4 letters. She chooses randomly the delivery order.**
    - **(a) What is the probability of choosing first the closest recipient?**
    - **(b) What is the probability of delivering a letter immediately before b letter?**

(a)

If we name the letters a, b, c, d, the possible delivery orderings are a, b, c, d / a, c, d, b / .... So the total number of outcomes is the number of permutations (different orderings) of those elements: $P_4 = 4! = 24$.

We can apply Laplace's rule, because all those 24 results are equiprobable, as the postman choose among them randomly.

Let's suppose the closest recipient is a. The outcomes resulting in the first recipient being the closest one are: a, b, c, d / a, c, b, d / a, d, b, c/ .... We can say then that a is fixed, so we can change the ordering of the other 3 elements (b,c,d). So the number of outcomes resulting in the event we are looking for is the number of permutations of 3 elements: $P_3 = 3! = 6$.

So, *P[the first recipient = the closest one]*$=\frac{6}{24} = 0.25$.

We can solve it in a much easier manner:

- if the delivery is random, the first recipient may be a, b, c or d, that is, we have 4 possible outcomes;
- but in the probability we are looking for, a (the closest recipient) must be the first one; and a is just one outcome.
- so, *P[the first recipient = the closest one]*$=\frac{1}{4} = 0.25$.

(b)

The total number of outcomes is the same (24), as we have the same situation (we're looking just for another probability).

Outcomes resulting in ab being consecutive are: $abcd$, $cdab$ and so on. So, ab is fixed and we can arrange it along with the other 2 letters (c and d). In total, we have 3 items (ab, c and d) that we can arrange in any order. The number of ways to do so is: $P_3 = 3! = 6$.

So: *P[deliver a-b in that order]*$=\frac{6}{24} = 0.25$.

14. **We have received 4 only one night bookings in our hostel for the next week, but we don't know the exact day. What is the probability of all of them being in different days?**

To apply Laplace's rule we have to count firts the number of all possible outcomes. As we don't know anything about the booked days, we must think they will happen randomly.

To list the possible outcomes let's make some coding:

Monday: 1 / Tuesday: 2 / .... / Sunday: 7.

Some we can express outcomes in this manner:

1st host	2nd host	3rd host	4th host
1	2	3	4
1	1	2	3
2	2	2	2
...	...	...	...

NUMBER OF ALL POSSIBLE OUTCOMES?

- Choosing problem
- $k = 4$
- $n = 7$ (from Monday to Sunday, 1 to 7)
- Repetition YES: 1123 is a possible outcome.
- Order (Y/N)?
- (Long, and important) explanation about order: if we don't take into account the order, 1234, 2341, 4321, and so on, are the same and we count these as they were just one. On the other hand, 2222, e.g., only happens with 2222. So, with order NO, 1234 and 222 outcomes would have different probabilities (1234 more probable) and we couldn't apply Laplace's rule. So we must take into the account the order, in order to take 1234, 1324, 1432, ..., and 2222 as different and EQUIPROBABLE outcomes.
- Hence, order YES.
- Then, $VR_{n=7}^{k=4} = 7^4$

OUTCOMES RESULTING IN DIFFERENT DAYS (that's the probability we want to calculate)? In this case we want to list all the outcomes with different days, that is to say, numbers from 1 to 7. It's a choosing problem, as the total number of outcomes, with $k = 4$ and $n = 7$, order YES, but now with NO repetition, as the days must be different. Hence, we must use the formula for variations to list those outcomes:

$$V_{n=7}^{k=4} = \frac{7!}{(7-4)!} = 7 \times 6 \times 5 \times 4$$

SO, THE PROBABILITY IS:

$$P[hosts\ in\ different\ days] = \frac{7 \times 6 \times 5 \times 4}{7^4} = 0.349$$

GENERAL RULE:
except for urn problems,
in order to apply Laplace's rule,

order is YES!

17. **60 people are attending a congress. 40 of them speak only Spanish, 20 of them French. What is the probability of two of them, randomly chosen, understand each other.**

Let's do coding:

- S: one person speaking Spanish
- F: one person speaking French

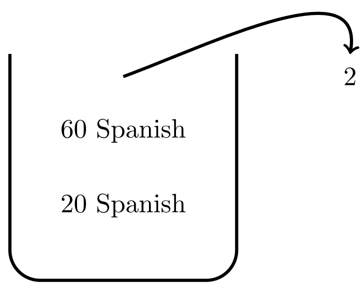
So the probability we want to calculate may be expressed in this way:

$$P[SS \cup FF]$$

When we have an union of events, first of all we must *always* ask ourselves if the events in the union are mutually exclusive. In this case they are, as there is no people speaking both French and Spanish; hence, the probability of the union is the addition of the probability of the events in the union:

$$P[SS \cup FF] = P[SS] + P[FF] = \frac{\binom{40}{2}}{\binom{60}{2}} + \frac{\binom{20}{2}}{\binom{60}{2}}$$

To calculate the latter simple probabilities we have taken the urn model as the basic reference:



30. **You have a 4 pieces batch, but you don't know which of them are faulty or faultless. Trying to improve the quality of the batch, you draw pieces one by one: if the piece you draw is faulty, you withdraw that piece and put into the batch 2 faultless pieces; if the piece you draw is faultless, you return it into the batch. You draw 3 pieces.**

    (a) **If you have 2 faulty and 2 faultless pieces in the batch, calculate the probabilities of all the possible outcomes, taking the order into account.**

    (b) **If the three pieces you have drawn are *oxo*, what should you deduce about the number of faulty pieces in the urn at the beginning?**

(a) The possible outcomes after drawing 3 pieces for a *ooxx* batch and their probabilities are:

Outcomes	Original batch	Batch after 1st piece	Batch after 2nd piece	Probability of the outcome
ooo	ooxx	ooxx	ooxx	$\frac{2}{4} \times \frac{2}{4} \times \frac{2}{4} = 0.125$
oox	ooxx	ooxx	ooxx	$\frac{2}{4} \times \frac{2}{4} \times \frac{2}{4} = 0.125$
oxo	ooxx	ooxx	ooxoo	$\frac{2}{4} \times \frac{2}{4} \times \frac{4}{5} = 0.200$
xoo	ooxx	ooxoo	ooxoo	$\frac{2}{4} \times \frac{4}{5} \times \frac{4}{5} = 0.320$
oxx	ooxx	ooxx	ooxoo	$\frac{2}{4} \times \frac{2}{4} \times \frac{1}{5} = 0.050$
xox	ooxx	ooxoo	ooxoo	$\frac{2}{4} \times \frac{4}{5} \times \frac{1}{5} = 0.080$
xxo	ooxx	ooxoo	oooooo	$\frac{2}{4} \times \frac{1}{5} \times \frac{6}{6} = 0.100$
xxx	ooxx	ooxoo	oooooo	$\frac{2}{4} \times \frac{1}{5} \times \frac{0}{6} = 0.000$
				1

All the outcomes and their probabilities give what we call a *probability distribution*. As it can be seen, all the probabilities sum to 1.

(b) Here we have the inverse problem: we don't know the composition of the batch, but we know the outcome of the 3 pieces we have drawn. We must conclude something about the composition of the box, based on that outcome. We may have 1, 2, 3 or 4 faulty pieces in the batch at the beginning. It's not possible to have 0 faultless items as the first piece drawn is *o* (faultless). The probabilities of being *oxo* after drawing 3 pieces from each of those possible batches, taking into account the changes along, are as follows:

Original batch	Batch after *o*	Batch after *ox*	Probability of *oxo*
oxxx	oxxx	oxxoo	$\frac{1}{4} \times \frac{3}{4} \times \frac{3}{5} = 0.1125$
ooxx	ooxx	ooxoo	$\frac{2}{4} \times \frac{2}{4} \times \frac{4}{5} = 0.4000$
ooox	ooox	ooooo	$\frac{3}{4} \times \frac{1}{4} \times \frac{5}{5} = 0.1875$
oooo	oooo	-	$\frac{4}{4} \times \frac{0}{4} = 0.0000$

We see now, but it could be seen previously, that it's not possible to have with that outcome a batch of type *oooo*.

Among the other probabilities, we seen that *oxo* outcome is the most probable with *ooxx* batch. So, we opt for that batch as the best choice given the *oxo* outcome.

We call these probabilities *verosimilities*, because they are about only one outcome by with different options (in this case, batches). Verosimilies are used to choose between different options about something we don't know, based on a given outcome. Notice the difference with the concept of *probability distribution*.

32. Possible daily sales at an auto dealer are 1 to 4, with 0.1 to 0.4 probabilities respectively, with total independence. What is the probability of selling 6 or more autos in two days?

$$P[\mathbf{X} \geq 6] = 0.08 + 0.09 + 0.12 + 0.08 + 0.12 + 0.16$$

33. A student takes two partial exams in a course. The probabilities of passing these exams are 0.5 and 0.6 respectively. Partial exams are saved and in the final exam the student must take only the failed partial exams. The probabilities of passing these final partial exams are 0.7 and 0.8 respectively. What is the probability of finally passing both exams?

Hence,

$$P[passing\ both\ exams] = 0.5 \times 0.6 + 0.5 \times 0.4 \times 0.8 + 0.5 \times 0.6 \times 0.7 + 0.5 \times 0.4 \times 0.7 \times 0.8$$

35. We have three applicants for a job. Two exams will be hold through the recruitment; the probability for passing the first exam is 0.6, and 0.3 for the second one. What is the probability of remaining 0, 1, 2 and 3 applicants after the exams?

Applicants after two exams	Probability
0	$0.4^3 + (0.4^2 \times 0.6 \times 3 \times 0.7) + (0.4 \times 0.6^2 \times 3 \times 0.7^2) + (0.6^3 \times 0.7^3)$
1	$(0.4^2 \times 0.6 \times 3 \times 0.3) + (0.4 \times 0.6^2 \times 3 \times 0.7 \times 0.3 \times 2) + (0.6^3 \times 0.7^2 \times 0.3 \times 3)$
2	$(0.4 \times 0.6^2 \times 3 \times 0.3^2) + (0.6^3 \times 0.7 \times 0.3^2 \times 3)$
3	$0.6^3 \times 0.3^3$
	Sum=1

36. Our stock price increases with a 0.4 probability, if it increased the day before; and decreases with a 0.7 probability, under the same circumstances. Through the next four days, what is the probability of the stock price increasing over two days and decreasing over the other two? Remark: last day stock price increased.

Hence,

$$P[2\ days\ increasing\ and\ 2\ days\ decreasing] =$$

$$0.4\times0.4\times0.6\times0.3+0.4\times0.6\times0.7\times0.6+0.4\times0.6\times0.3\times0.7+0.6\times0.7\times0.4\times0.6+0.6\times0.7\times0.6\times0.7+0.6\times0.3\times0.7\times0.4$$

41. Broadly, the customers of a new product may be 10%, 20%, 30% or 40% of the total customers in the market. There is total uncertainty about the probability of those values. A survey took 10 people and among them 6 claimed they would purchase that product. How should we change the a priori probabilites about the percentage of customers?

As we have total uncertainty about the percentage of customers, we assign to all percentages the same a priori probability: 0.25.

B: 6 consumers among 10 people

Verosimilities are thought in this manner: if percentage of customers is 10%, probability for a given person of being a customer is 0.1. For 10 people we have just to multiply the probabilities, and as any ordering is not set we multiply also by the permutations.

Following calculations are straightforward.

A_i	$P(A_i)$	$P(B/A_i)$	$P(A_i) \times P(B/A_i)$	$P(A_i/B)$
consumer percentage: 10%	**0.25**	$0.1^6 \times 0.9^4 \times \frac{10!}{6!4!} = 0.00013$	**0.00003**	**0.0007**
consumer percentage: 20%	**0.25**	$0.2^6 \times 0.8^4 \times \frac{10!}{6!4!} = 0.00550$	**0.00137**	**0.035**
consumer percentage: 30%	**0.25**	$0.3^6 \times 0.7^4 \times \frac{10!}{6!4!} = 0.03675$	**0.00919**	**0.239**
consumer percentage: 40%	**0.25**	$0.4^6 \times 0.6^4 \times \frac{10!}{6!4!} = 0.11147$	**0.02787**	**0.724**
	1		**0.03846**	**1**

The result are intuitive: survey tells us the consumer are 60%. The closest percentage among those we take into account as hypothesus is 40%. Hence, that will be the winner,the percentage we should bet for as the true one.

---

42. In the customer database of an insurance company, we have these data: 100 policy holder had an accident last year, 200 policy holder didn't have any accident last year, 400 policy holders didn't have any accident last two years, and 500 policy holders didn't have any accident last three years. According to our estimations, a person that had accident last year has a 0.22 probability of having another accident next year, and that probability declines by 0.04 by year without accident. We have just received a claim about an accident, but we know nothing about the policy holder. Into which policy holder group should we enter him?

B: customer claimed a recent accident

A_i	$P(A_i)$	$P(B/A_i)$	$P(A_i) \times P(B/A_i)$	$P(A_i/B)$
had accident last year	$\frac{100}{1200}$	**0.22**		
hadn't accident last year	$\frac{200}{1200}$	**0.18**		
hadn't accident last 2 years	$\frac{400}{1200}$	**0.14**		
hadn't accident last 3 years	$\frac{500}{1200}$	**0.10**		
	1			**1**

The policy holder should be in the group with the biggest a posteriori probability. Do the calculations!

[43.] We have 6 faulty washing machines in stock, and 12 faultless. We sold 8 of them. As we don't know which of these are faulty and faultless, we delivered them randomly. We expect the customer who takes a faulty machine will claim.

i. Give the probability mass function and the cumulative distribution function of the number of claims.

ii. We need a spare part for each claim. How many spare parts do we need in order to satisfy all the claims surely.

iii. And in order to satisfy all the claims with a probability of 0.8?

Notice that the number of spare parts to be required is the same as the number of claims and the latter the same as the number of faulty machines:

x	$P[X = x]$	$F(x) = P[X \leq x]$
0	$\frac{\binom{6}{0}\binom{12}{6}}{\binom{18}{8}} = 0.011$	**0.011**
1	$\frac{\binom{6}{1}\binom{12}{5}}{\binom{18}{8}} = 0.109$	**0.120**
2	$\frac{\binom{6}{2}\binom{12}{4}}{\binom{18}{8}} = 0.317$	**0.437**
3	$\frac{\binom{6}{3}\binom{12}{3}}{\binom{18}{8}} = 0.362$	**0.799**
4	$\frac{\binom{6}{4}\binom{12}{2}}{\binom{18}{8}} = 0.170$	**0.969**
5	$\frac{\binom{6}{5}\binom{12}{1}}{\binom{18}{8}} = 0.030$	**0.999**
6	$\frac{\binom{6}{6}\binom{12}{0}}{\binom{18}{8}} = 0.001$	**1**

1

- **First column is the random variable.**
- $P[X = x]$ **is the probability mass function.**
- $F(x) = P[X \leq x]$ **is the distribution function.**
- **Notice that the maximum number of faulty machines is not 8 (the total number of machines sold), but 6, because totally we have not more than 6 faulty machines.**
- **The number of spare parts needed to cover all the claims surely is 6, because is not possible to have more than 6 faulty machines among the machines we have sold.**
- **As for all *how many* type problems with a given probability, the best thing is to take a given value and test it, and on that basis check for the exact value we are seeking for. So let's take, e.g., 4 spare parts: they will be enough if the number of claims is 4 or less: the probability is 0.969. We reach (by far) the 0.8 probability (notice that 0.8 probability must be taken as a minimum and not as an exact value. Let's see if it's possible with lesser parts to keep the probability bigger than 0.8. Reducing to 3 parts, the probability to cover all claims is 0.799, so we don't reach (narrowly) the intended probability. So the solution is 4 spare parts.**

[44.] The number of visits until we get a new customer in a web page is a r.v. given by this function:

$$P[X = x] = 0.1 + \frac{k}{x} \ ; \ x = 1, 2, 3, 4$$

i. Calculate k, $P[X = x]$ to be a mass function.

ii. Give the mass function and the cumulative distribution function as a table.

(a) As sum of probabilities must be 1:

$$P[X = 1] + P[X = 2] + P[X = 3] + P[X = 4] = 0.1 + \frac{k}{1} + 0.1 + \frac{k}{2} + 0.1 + \frac{k}{3} + 0.1 + \frac{k}{4} = 0.4 + \frac{25k}{12} = 1 \to k = 0.288$$

(b)

x	$P[X = x]$	$F(x) = P[X \leq x]$
1	$0.1 + \frac{0.288}{1} = 0.388$	0.288
2	$0.1 + \frac{0.288}{2} = 0.244$	0.632
3	$0.1 + \frac{0.096}{2} = 0.196$	0.828
4	$0.1 + \frac{0.288}{4} = 0.172$	1.000

Sum=1

- **The first column is the random variable.**
- **The second column is the mass function.**
- **The third column is the distribution function.**
- **The whole table would be the probability distribution.**

[45.] The distribution of the X random variable is given by:

$$P[X = x] = \frac{1}{k}\ ;\ x = 1, 2, \ldots, k$$

Calculate k to be a mass function.

Sum of probabilities must be 1:

$$P[X = 1] + P[X = 2] + \cdots + P[X = k] = \underbrace{\frac{1}{k} + \frac{1}{k} + \cdots + \frac{1}{k}}_{\textbf{k times}} = k \times \frac{1}{k} = 1$$

As the equality holds for every k, we say k is a parameter. The concept of parameter is very important in statistics, as the point is to estimate their value or test a given value for them: we do that from data and using statistical inference techniques.

[49.] A random variable follows this distribution:

$$f(x) = k - x \; ; \; 0 < x < k$$

i. Calculate k the previous function to be a pdf.

ii. Give the cdf.

iii. Calculate $P[0.5 < X < 1]$, by means of both the pdf and the cdf.

---

i. Area under pdf must be 1:

$$\int_0^k (k-x)dx = \left[kx - \frac{x^2}{2}\right]_0^k = \left[k^2 - \frac{k^2}{2}\right] - \left[k \times 0 - \frac{0^2}{2}\right] = \frac{k^2}{2} = 1 \rightarrow k = \sqrt{2} = 1.41$$

Hence, $f(x) = 1.41 - x \; ; \; 0 < x < 1.41$**.**

ii. $F(x) = P[X < x] = \int_{inf}^{x} f(x)dx = \int_0^x (1.41 - x)dx = 1.41x - \frac{x^2}{2} \; ; \; 0 \leq x \leq 1.41$

iii. $P[0.5 < X < 1]$**?**

- **by the pdf:**

$$\int_{0.5}^{1} (1.41 - x)dx = \left[1.41x - \frac{x^2}{2}\right]_{0.5}^{1} = 0.33$$

- **by the cdf:**

$$P(X < 1) - P(X < 0.5) = F(x = 1) - F(x = 0.5) = \left(1.41 \times 1 - \frac{1^2}{2}\right) - \left(1.41 \times 0.5 \times 1 - \frac{0.5^2}{2}\right) = 0.33$$

[50.] Daily production (kg) is rv, depending on the number of machines (k):

$$f(x) = \frac{2x}{100k^2} \ ; \ 0 < x < 10k$$

(a) **Proof that k is a parameter.**

(b) **With 4 machines, what is the probability of production exceeding 30?**

(c) **Give cdf, depending on k, and based on that, proof k is a parameter.**

(d) **If product is packed by the kilogram, how many packs do we need, with 6 machines, in order to be the probability of packing all the production 0.8? And to be 0.9?**

(e) **How many machines do we need in order to the probability of the production being more than 60 reaching 0.5?**

(a) **We have just to proof that the condition for $f(x)$ being a density function holds for every k:**

$$\int_\Omega f(x)dx = \int_0^{10k} \frac{2x}{100k^2}dx = \left[\frac{2x^2}{200k^2}\right]_0^{10k} = \left[\frac{200k^2}{200k^2}\right] - [0] = 1$$

(b) **We have now $k = 4$. So the density function will be:**

$$f(x) = \frac{2x}{1600} \ ; \ 0 < x < 40$$

And now the probability of production exceeding 30:

$$P[X > 30] = \int_{30}^{40} \frac{2x}{1600}dx = \left[\frac{2x^2}{3200}\right]_{30}^{40} = \left[\frac{3200}{3200}\right] - \left[\frac{1800}{3200}\right] = 0.4375$$

(c)

$$F(x) = \int_{inf}^{x} \frac{2x}{100k^2}dx = \left[\frac{2x^2}{200k^2}\right]_0^x = \left[\frac{x^2}{100k^2}\right]_0^x = \left[\frac{x^2}{100k^2}\right] - [0] = \frac{x^2}{100k^2} \ ; \ 0 \leq x \leq 10k$$

To proof k is a parameter, we apply the conditions:

- $F(inf) = 0 \rightarrow F(0) = 0$ **(it's not related cto k, so it doesn't tell us anything about it.**
- $F(sup) = 1 \rightarrow F(10k) = \frac{(10k)^2}{100k^2} = \frac{100k^2}{100k^2} = 1$. **The condition holds for every k, so k is a parameter.**

(d) **We have $k = 6$.**

It's a *how many* type problem; so let's test a given value. As with the maximum number of packages is 40 (as the maximum production is 40 kg), let's test 30 packages. 30 packages will be enough when the production is under 30:

$$P[enough] = P[X < 30] = \int_0^{30} \frac{2x}{3600}dx = \left[\frac{2x^2}{7200}\right]_0^{30} = \left[\frac{1800}{7200}\right] - [0] = 0.25$$

As we don't reach 0.8, it's not enough. So to calculate the exact value we set the probability to 0.8 and the number of packages to p:

$$\left[\frac{2x^2}{7200}\right]_0^p = \left[\frac{2p^2}{7200}\right] - [0] = 0.8 \rightarrow p = 53.66$$

As the packages are by kilo, and 0.8 is a minimum probability, we must take 54 packages.

To be the probability 0.9, we will need more packages:

$$\left[\frac{2x^2}{7200}\right]_0^p = \left[\frac{2p^2}{7200}\right] - [0] = 0.9 \rightarrow p = 56.92$$

And so we will need 57 packages.

[51.] The time to complete a task is given by the following cdf:

$$F(x) = \frac{x^3 - 8}{19} \;;\quad 2 \leq x \leq 3$$

i. Proof it's a cdf.

ii. Give the pdf.

iii. Calculate $P[X \geq 2.5]$ **by means of both the pdf and the cdf.**

---

i. Three conditions must be held:

- $F(inf) = 0 \rightarrow F(x=2) = 0$ **(ok)**
- $F(sup) = 1 \rightarrow F(x=3) = 1$ **(ok)**
- $gorakorra \rightarrow F'(x) = \frac{3x^2}{19} > 0, \forall x \in [2,3]$ **(ok)**

ii. We get the pdf derivating the cdf:

$f(x) = F'(x) = \frac{3x^2}{19} \;;\quad 2 < x < 3$

iii. As the probability in a point is 0, we don't take into account the equality.

Using the cdf: $P[X > 2.5] = 1 - P[X < 2.5] = 1 - F(x = 2.5) = 0.59$

Using the pdf: $P[X > 2.5] = \int_{2.5}^{3} \frac{3x^2}{19} dx = \left(\frac{x^3}{19}\right)_{2.5}^{3} = \left(\frac{3^3}{19}\right) - \left(\frac{2.5^3}{19}\right) = 0.59$

[53.] The weight of an *clementina* orange follows this distribution:

$$f(x) = \frac{2}{10000}(x-100)\ ;\ 100 < x < 200$$

Calculate the probability of the weight being exactly 110 gr, both in a theoretical way and a practical way, known than balance gives 100-110-120-130-... weights. And 130 gr or less?

Some examples:

True weight	Balance weight
108	110
101	100
112	110

So:

$$P[X=110] = P[105 < X < 115] = \int_{105}^{115} \frac{2}{10000}(x-100)dx$$

Now we calculate $P[X \leq 130]$**?**

It's not $F(x=130)$**, as including 130 is relevant, in spite of using a continuous distribution. So:**

$$P[X \leq 130] = P[X < 135] = F(x=135)$$

[54.]

The number of customers entering a store follows this distribution:

$$f(x) = \frac{1}{100}\ ; \quad 200 < x < 300$$

What is the probability of entering exactly 250 customers? And 250-260 customers (closed interval)?

In the discrete domain numbers of persons is like this: ..., 248, 249, 250, 251, 252, ...

$$P[X = 250] = P[249.5 < X < 250.5] = \int_{249.5}^{250.5} \frac{1}{100} dx$$

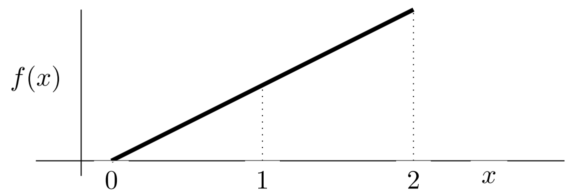
And for the [250-260] closed interval, including 250 and 260 is relevant so:

$$P[250 \leq X \leq 260] = P[249.5 < X < 260.5] = \int_{249.5}^{260.5} \frac{1}{100} dx$$

[57.] The number of items produced by a machine (thousands) distributes in this way:

$$f(x) = \frac{x}{2}; 0 < x < 2$$

Plot the density function and approximate the expectation at first sight. Calculate it.



From $x = 1$ **(the middle point) there is bigger probability density (it's more probable) upwards (**$x > 1$**) than downwards (**$x < 1$**). Hence, the expected value will be bigger than 1.**

$$\mu = E[x] = \int_{\Omega} x f(x) dx = \int_0^2 x \frac{x}{2} dx = \int_0^2 \frac{x^2}{2} dx = \left[\frac{x^3}{6} \right]_0^2 = \left[\frac{2^3}{6} \right] - \left[\frac{0^3}{6} \right] = \frac{8}{6} = 1.333 = 1333 \; items$$

[59.] The tenperature in a freezer in any moment has the following distribution:

$$f(x) = \frac{1}{a} \ ; \ 0 < x < a$$

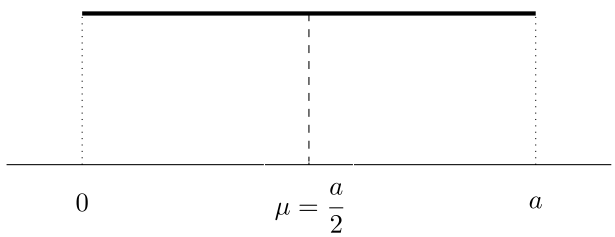
i. Plot the distribution and interpret it about the mean value.

ii. Calculate the mean value and the standard deviation.

iii. Calculate the 3rd central moment.

(a)

As probability is uniform or balanced along the support, the expected value will be in the middle point:



(b)

Expected value:

$$\mu = E[X] = \alpha_1 = \int_\Omega x f(x)dx = \int_0^a x\frac{1}{a}dx = \left[\frac{x^2}{2a}\right]_0^a = \left[\frac{a^2}{2a}\right] - [0] = \frac{a}{2}$$

To calculate the standard deviation we have to calculate the variance:

$$\sigma_X^2 = \alpha_2 - \alpha_1^2$$

Second order moment about the origin:

$$\alpha_2 = E[X^2] = \int_\Omega x^2 f(x)dx = \int_0^a x^2\frac{1}{a}dx = \left[\frac{x^3}{3a}\right]_0^a = \left[\frac{a^3}{3a}\right] - [0] = \frac{a^2}{3}$$

Variance:

$$\sigma_X^2 = \alpha_2 - \alpha_1^2 = \frac{a^2}{3} - \left(\frac{a}{2}\right)^2 = \frac{a^2}{12}$$

Standard deviation:

$$\sigma_X = \sqrt{\frac{a^2}{12}}$$

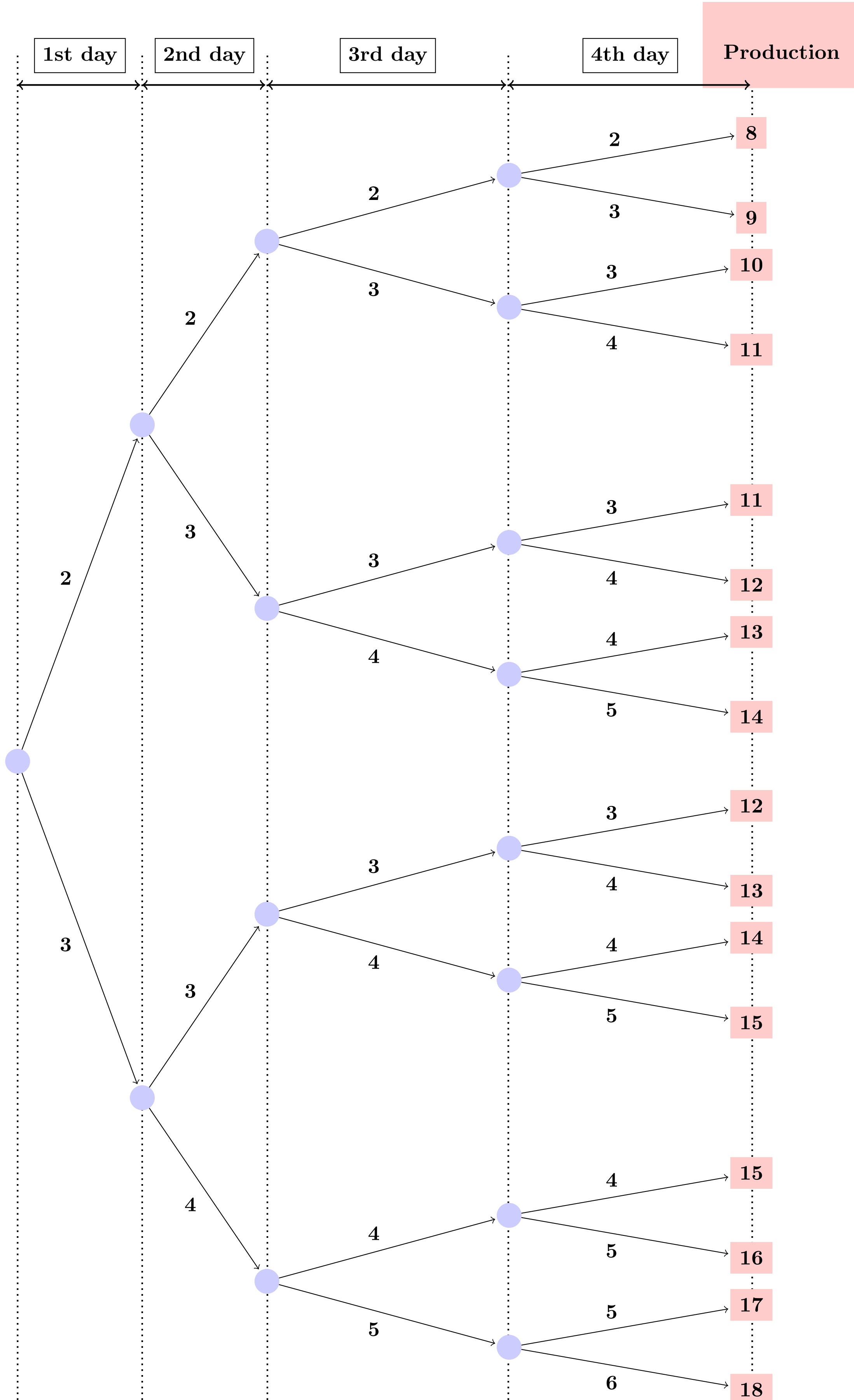
(c)

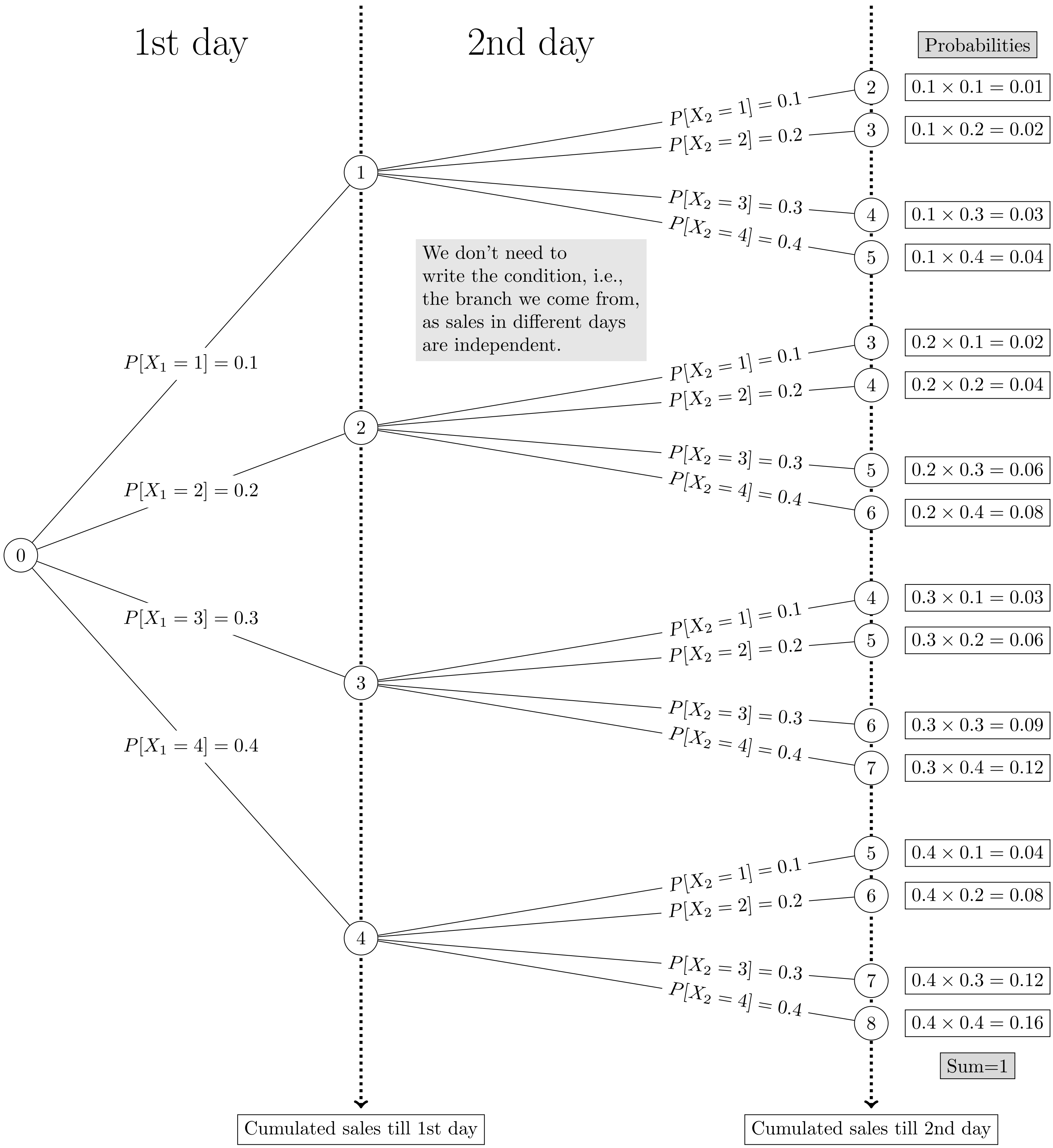
Third central moment:

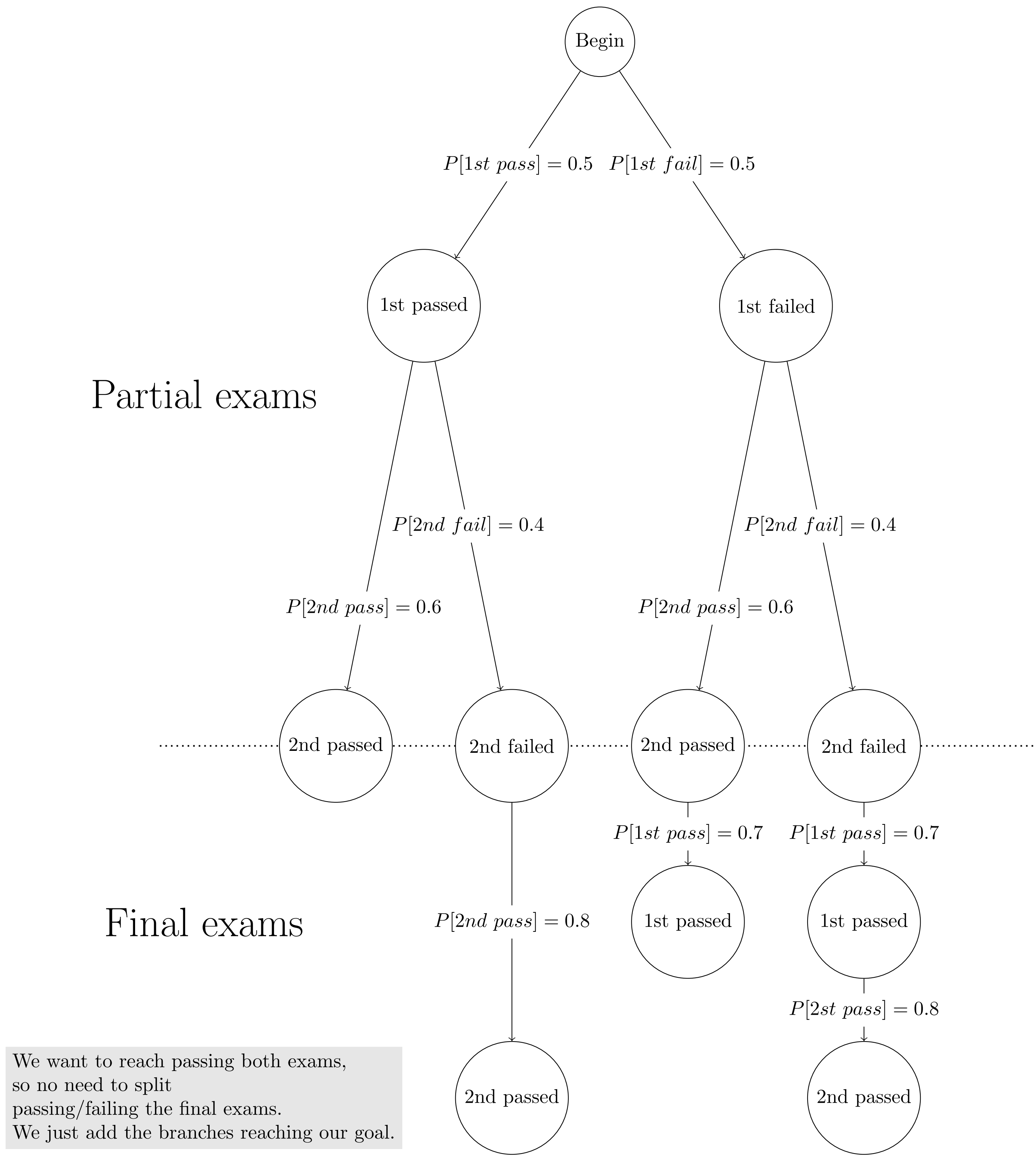
$$\alpha_2 = \int_\Omega (x-\mu)^3 f(x)dx = \int_0^a \left(x - \frac{a}{2}\right)^3 \frac{1}{a}dx = \left[\frac{1}{a}\frac{(x-\frac{a}{2})^4}{4}\right]_0^a = \left[\frac{1}{a}\frac{(a-\frac{a}{2})^4}{4}\right] - \left[\frac{1}{a}\frac{(0-\frac{a}{2})^4}{4}\right] = 0$$

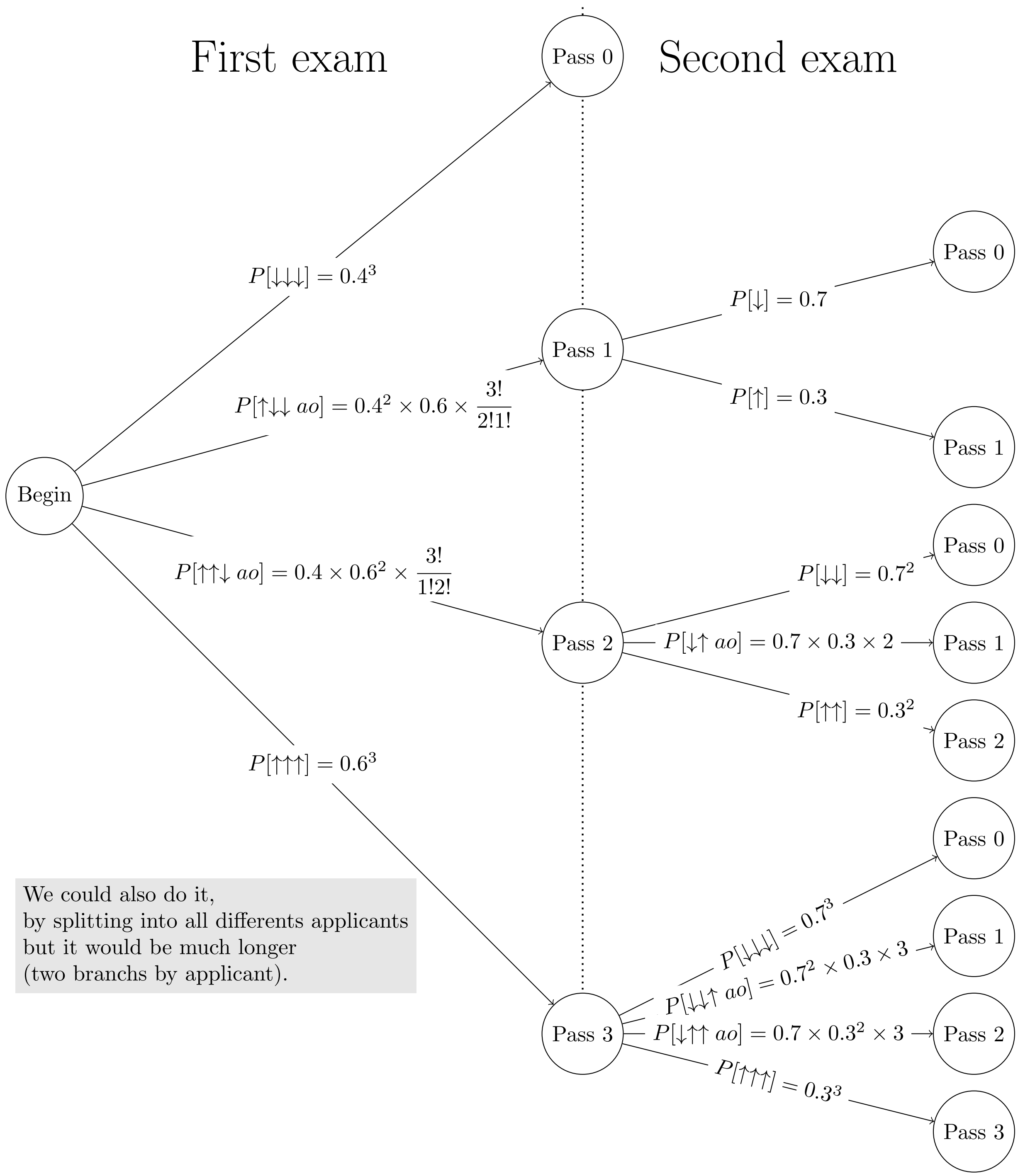
[61.]

As a tree:

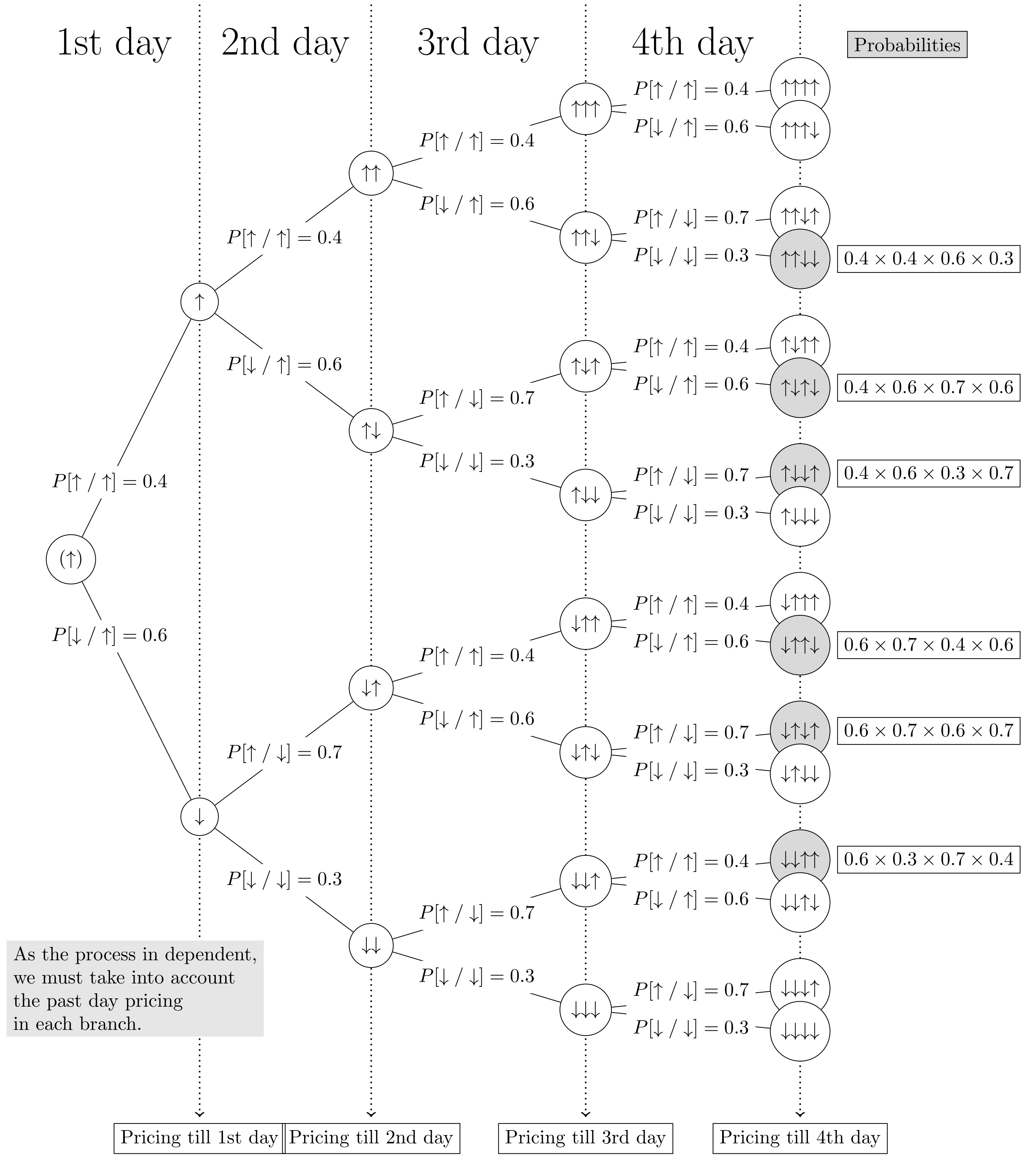

1st day
2nd day
3rd day
4th day
Production
2
3
2
3
3
4
2
3
3
4
2
3
3
4
4
5
3
4
3
4
4
5
4
5
4
5
5
6
8
9
10
11
11
12
13
14
12
13
14
15
15
16
17
18

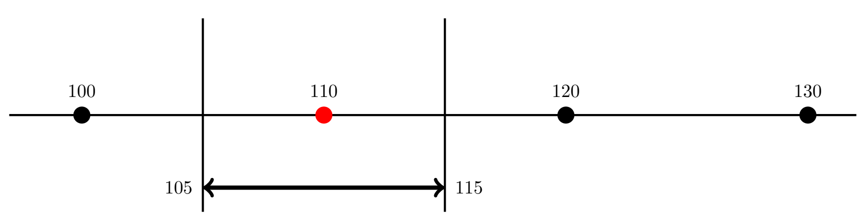

1st day
2nd day
Probabilities
0
P[X1 = 1] = 0.1
P[X1 = 2] = 0.2
P[X1 = 3] = 0.3
P[X1 = 4] = 0.4
1
2
3
4
P[X2 = 1] = 0.1
P[X2 = 2] = 0.2
P[X2 = 3] = 0.3
P[X2 = 4] = 0.4
We don't need to write the condition, i.e., the branch we come from, as sales in different days are independent.
0.1 × 0.1 = 0.01
0.1 × 0.2 = 0.02
0.1 × 0.3 = 0.03
0.1 × 0.4 = 0.04
0.2 × 0.1 = 0.02
0.2 × 0.2 = 0.04
0.2 × 0.3 = 0.06
0.2 × 0.4 = 0.08
0.3 × 0.1 = 0.03
0.3 × 0.2 = 0.06
0.3 × 0.3 = 0.09
0.3 × 0.4 = 0.12
0.4 × 0.1 = 0.04
0.4 × 0.2 = 0.08
0.4 × 0.3 = 0.12
0.4 × 0.4 = 0.16
Sum=1
Cumulated sales till 1st day
Cumulated sales till 2nd day


Begin
P[1st pass] = 0.5
P[1st fail] = 0.5
1st passed
1st failed
Partial exams
P[2nd fail] = 0.4
P[2nd pass] = 0.6
P[2nd fail] = 0.4
P[2nd pass] = 0.6
2nd passed
2nd failed
2nd passed
2nd failed
P[1st pass] = 0.7
P[1st pass] = 0.7
Final exams
P[2nd pass] = 0.8
1st passed
1st passed
P[2st pass] = 0.8
We want to reach passing both exams,
so no need to split
passing/failing the final exams.
We just add the branches reaching our goal.
2nd passed
2nd passed



We could also do it,
by splitting into all differents applicants
but it would be much longer
(two branchs by applicant).


1st day
2nd day
3rd day
4th day
Probabilities
(↑)
P[↑ / ↑] = 0.4
P[↓ / ↑] = 0.6
↑
↓
P[↑ / ↑] = 0.4
P[↓ / ↑] = 0.6
P[↑ / ↓] = 0.7
P[↓ / ↓] = 0.3
↑↑
↑↓
↓↑
↓↓
P[↑ / ↑] = 0.4
P[↓ / ↑] = 0.6
P[↑ / ↓] = 0.7
P[↓ / ↓] = 0.3
P[↑ / ↑] = 0.4
P[↓ / ↑] = 0.6
P[↑ / ↓] = 0.7
P[↓ / ↓] = 0.3
↑↑↑
↑↑↓
↑↓↑
↑↓↓
↓↑↑
↓↑↓
↓↓↑
↓↓↓
P[↑ / ↑] = 0.4
P[↓ / ↑] = 0.6
P[↑ / ↓] = 0.7
P[↓ / ↓] = 0.3
↑↑↑↑
↑↑↑↓
↑↑↓↑
↑↑↓↓
0.4 × 0.4 × 0.6 × 0.3
↑↓↑↑
↑↓↑↓
0.4 × 0.6 × 0.7 × 0.6
↑↓↓↑
0.4 × 0.6 × 0.3 × 0.7
↑↓↓↓
↓↑↑↑
↓↑↑↓
0.6 × 0.7 × 0.4 × 0.6
↓↑↓↑
0.6 × 0.7 × 0.6 × 0.7
↓↑↓↓
↓↓↑↑
0.6 × 0.3 × 0.7 × 0.4
↓↓↑↓
↓↓↓↑
↓↓↓↓
As the process in dependent, we must take into account the past day pricing in each branch.
Pricing till 1st day
Pricing till 2nd day
Pricing till 3rd day
Pricing till 4th day


100
110
120
130
105
115

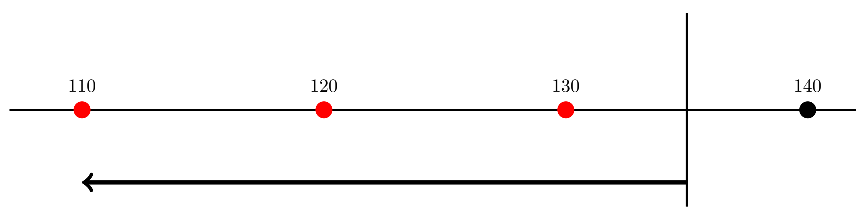


Figure 1: $Explanation : P[X \leq 130]$.

1st day	2nd day	3rd day	4th day	Prod (x)	p(x)	xp(x)	Time for 8 items (days)
2 (0.3)	2 (0.3)	2(0.3)	2 (0.3)	8	0.0081	0.0648	4
2 (0.3)	2 (0.3)	2 (0.3)	3 (0.7)	9	0.0189	0.1701	3.66
2 (0.3)	2 (0.3)	3 (0.7)	3 (0.3)	10	0.0189	0.1890	3.33
2 (0.3)	2 (0.3)	3 (0.7)	4 (0.7)	11	0.0441	0.4851	3.25
2 (0.3)	3 (0.7)	3 (0.3)	3 (0.3)	11	0.0189	0.2079	3
2 (0.3)	3 (0.7)	3 (0.3)	4 (0.7)	12	0.0441	0.5292	3
2 (0.3)	3 (0.7)	4 (0.7)	4 (0.3)	13	0.0441	0.5733	2.75
2 (0.3)	3 (0.7)	4 (0.7)	5 (0.7)	14	0.1029	1.4406	2.75
3 (0.7)	3 (0.3)	3 (0.3)	3 (0.3)	12	0.0189	0.2268	2.66
3 (0.7)	3 (0.3)	3 (0.3)	4 (0.7)	13	0.0441	0.5733	2.66
3 (0.7)	3 (0.3)	4 (0.7)	4 (0.3)	14	0.0441	0.6174	2.5
3 (0.7)	3 (0.3)	4 (0.7)	5 (0.7)	15	0.1029	1.5435	2.5
3 (0.7)	4 (0.7)	4 (0.3)	4 (0.7)	15	0.0441	0.6615	2.25
3 (0.7)	4 (0.7)	4 (0.3)	5 (0.7)	16	0.1029	1.6464	2.25
3 (0.7)	4 (0.7)	5 (0.7)	5 (0.3)	17	0.1029	1.7493	2.20
3 (0.7)	4 (0.7)	5 (0.7)	6 (0.7)	18	0.2401	4.3218	2.20
					1	$\mu = 15$	

2-2-2-3 sequence: time for 8 pieces

Answer: 3 days + 2/3=0.66 days= 3.66 days.

Merging time values:

Days (y)	p(y)	yp(y)
2.20	**0.343**	**0.7546**
2.25	**0.147**	**0.33075**
2.5	**0.147**	**0.3675**
2.66	**0.063**	**0.16758**
2.75	**0.147**	**0.40425**
3	**0.0189+0.0441=0.063**	**0.189**
3.25	**0.0441**	**0.143325**
3.33	**0.0189**	**0.062937**
3.66	**0.0189**	**0.069174**
4	**0.0081**	**0.0324**
	1	$\mu = 2.521516$

[62.]

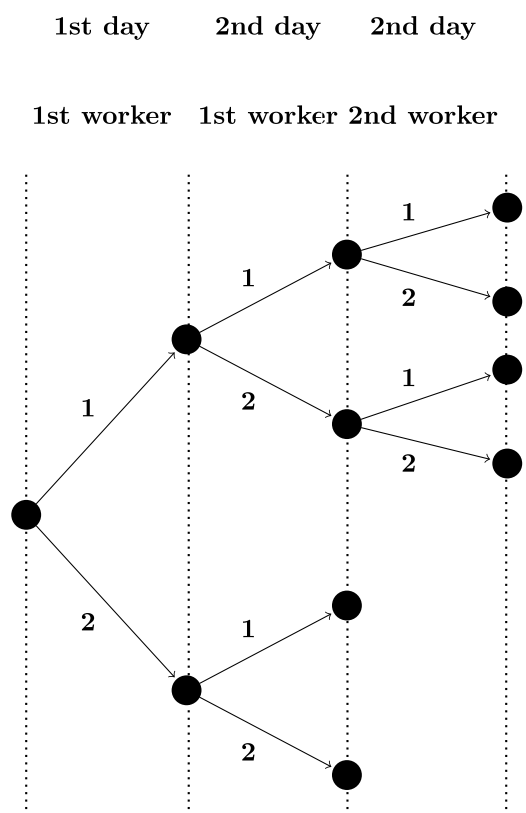
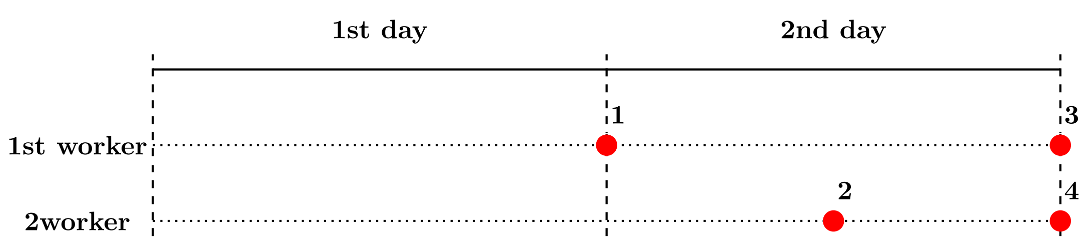


Table 1: **: Tree as a table.**

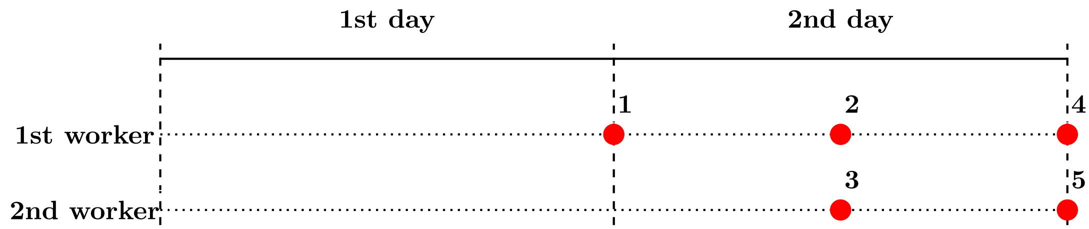
1st day 1st worker	2nd day 1st worker	2nd day 2 worker	prod(x)	p(x)=p(y)	Time till 2 items (y)
1 (0.4)	**1 (0.4)**	**1 (0.4)**	**3**	**0.064**	**2**

1-1-2 sequence: time till 2nd item



Answer: 1.5 days for 2nd item.

1-2-2 sequence: time till 2nd item



Answer: 1.5 days till the 2nd item.

Table 2: **Compact tables for total prod. and times till 2nd item**

x (prod)	p(x) probability	xp(x)	y (time)	p(y) probability	yp(y)

63

Table 3: : **Tables for μ and variance**

x	$p_A(x)$	$x_Ap(x)$	$x_A^2p(x)$	$p_B(x)$	$x_Bp(x)$	$x_B^2p(x)$	$p_C(x)$	$x_Cp(x)$	$x_C^2p(x)$	$p_D(x)$	$x_Dp(x)$	$x_D^2p(x)$
-2	**0.05**			**0**			**0**			**0.05**		
-1	**0.25**			**0.20**			**0.15**			**0.10**		
0	**0.30**			**0.50**			**0.40**			**0.35**		
1	**0.20**			**0.25**			**0.30**			**0.40**		
2	**0.20**			**0.05**			**0.15**			**0.10**		
	1	μ_A=	α_A^2=		μ_B=	α_B^2=		μ_C=	α_C^2=		μ_D=	α_D^2=

Table 4: μ, σ, $p(galdu)$: summary table

Choice	μ	σ	$p(losing)$
A			
B			
C			
D			

Table 5: Preferences two-by-two.

	A	B	C	D
A	-			
B		-		
C			-	
D				-

Table 6: μ, σ, $p(losing)$: preferences.

μ	σ	$p(losing)$

[64.] We have to choose betwen two stock options, whose gains distribute in the following way:

- **A stock option:**

$$f(x) = \frac{1}{2} - \frac{x}{8} \ ; \ 0 < x < 4$$

- **B stock option:**

$$f(x) = \frac{1}{5}; \ 0 < x < 5$$

If needed, you should use this utility function:

$$U = \frac{\mu}{\sigma}$$

Discuss which is the best choice, in the long term as in the short term.

In the long term we have to look only at the expectation. Here we calculate the expectation for A stock:

$$\mu_A = \int_0^4 x\left(\frac{1}{2} - \frac{x}{8}\right)dx = \left[\frac{x^2}{4} - \frac{x^3}{24}\right]_0^4 = 1.33$$

For B stock:

$$\mu_B = \int_0^5 x\frac{1}{5}dx = \left[\frac{x^2}{10}\right]_0^5 = 2.5$$

As it has a bigger expectation, in the long term we prefer B stock.

In the short term we have to loo at the expectation and the risk (measured by means of the variance). To calculate the variance, we calculate the 2nd raw moment:

$$\alpha_{2A} = \int_0^4 x^2\left(\frac{1}{2} - \frac{x}{8}\right)dx = \left[\frac{x^3}{6} - \frac{x^4}{32}\right]_0^4 = 10,66 - 8 = 2,66$$

And so variance is:

$$\sigma_A^2 = \alpha_{2A} - \alpha_{1A}^2 = 2.66 - 1.33^2 = 0.8911$$

We calculate in the same manner for B :

$$\alpha_{2B} = \int_0^5 x^2\frac{1}{5}dx = \left[\frac{x^3}{15}\right]_0^5 = 8.33$$

$$\sigma_B^2 = \alpha_{2B} - \alpha_{1B}^2 = 8.33 - 2.5^2 = 2.08$$

According to expectation B is better. According to variance A is better. To solve the dilemma we calculate the utility for both stocks:

$$U_A = \frac{\mu_A}{\sigma_A} = \frac{1.33}{\sqrt{0.8911}} = 1.41$$

$$U_B = \frac{\mu_B}{\sigma_B} = \frac{2.5}{\sqrt{2.08}} = 1.73$$

So according to that utility function B stock would be better in the short term.

[74] In a factory, we must produce 12 items for a customer. The probability of each item being faulty is 0.12. (a) Give the distribution of the faulty items among the 12 items.

$X \sim B(n = 12, p = 0.12)$**, assuming yhere is independence among different items (that's not very realistic: consecutive items faulty implies nest one being faulty with bigger probability).**

(b) Give the distribution of the faultless items among the 12 items.

$X \sim B(n = 12, p = 0.88)$

(b) Give the distribution of the faultless items among the 12 items.

$Y \sim B(n = 12, p = 0.88)$

(c) What is the probability of having 4 faulty items?

$P[X = 4] = 0.12^4 \cdot 0.88^8 \cdot \frac{12!}{4!8!}$

(d) What is the probability of having 3 faultless items?

$P[Y = 3] = 0.88^3 \cdot 0.12^9 \cdot \frac{12!}{3!9!}$

(e) What is the probability of having 3 faultless items and 9 faulty items?

Having 3 faultless items implies having 9 faulty items. So it will the same probability as that of the previous section.

(f) What is the probability of having 2 faulty items or less?

$P[X \leq 2] = P[X = 0] + P[X = 1] + P[X = 2] = 0.12^0 \cdot 0.88^{12} \cdot \frac{12!}{0!12!} + 0.12^1 \cdot 0.88^{11} \cdot \frac{12!}{1!11!} + 0.12^2 \cdot 0.88^{10} \cdot \frac{12!}{2!10!}$

(g) What is the probability of having less than 3 faulty items?

The same as the that of the previous section.

(h) What is the probability of having more than 8 faulty items?

$P[X > 8] = P[X = 9] + \cdots + P[X = 12] = 0.12^9 \cdot 0.88^3 \cdot \frac{12!}{9!3!} + \cdots + 0.12^{12} \cdot 0.88^0 \cdot \frac{12!}{12!0!}$

(i) Solve the last three questions by means of the Larson's nomogram.

Explanation given by the teacher.

(j) What is the probability of having 10 faulty items or more?

$P[X \geq 10] = P[X = 10] + \cdots + P[X = 12] = 0.12^{10} \cdot 0.88^2 \cdot \frac{12!}{10!2!} + \cdots + 0.12^{12} \cdot 0.88^0 \cdot \frac{12!}{12!0!}$

(k) What is the probability of the number of faulty events being between 4 and 6, both included?

$P4 \leq X \leq 6] = P[X = 4] + P[X = 5] + P[X = 6]$

And we apply for the each of them the known formula:

$P[X = x] = p^x \cdot (1 - p)^{(12-x)} \cdot \frac{12!}{x!(12 - x)!}$

(l) Solve (c)-(k) questions with R software.

```
> dbinom(4,12,0.12) #P[X=4];X~B(12,0.12)
[1] 0.03691404
> dbinom(3,12,0.88) #P[Y=3];Y~B(12,0.88)
[1] 7.735741e-07
> pbinom(2,12,0.12) #P[X<=2];X~B(12,0.12)
[1] 0.8332749
> pbinom(8,12,0.12,lower.tail=FALSE) #P[X> 8];X~B(12,0.12)
[1] 8.060138e-07
> 1-pbinom(8,12,0.12) #P[X> 8];X~B(12,0.12) (another way)
[1] 8.060138e-07
> x=9:12
> sum(dbinom(x,12,0.12)) #P[X> 8];X~B(12,0.12) (another way)
[1] 8.060138e-07
> pbinom(9,12,0.12,lower.tail=FALSE) #P[X>10];X~B(12,0.12)
[1] 3.243975e-08
> 1-pbinom(9,12,0.12) #P[X>10];X~B(12,0.12) (another way)
[1] 3.243975e-08
> pbinom(6,12,0.12)-pbinom(3,12,0.12) #P[4<=X<=6];X~B(12,0.12)
[1] 0.04624932
> x=4:6
> sum(dbinom(x,12,0.12)) #P[4<=X<=6];X~B(12,0.12) (another way)
[1] 0.04624932
```

(m) Which is the mean number of faulty items?

$\mu = E[X] = np = 12 \times 0.12 = 1.44$ **faulty items.**

(n) Solve by R: which is the most probable number of faulty items?

```
> x=0:12
> dbinom(x,12,0.12)
 [1] 2.156712e-01 3.529164e-01 2.646873e-01 1.203124e-01 3.691404e-02
 [6] 8.053972e-03 1.281314e-03 1.497639e-04 1.276397e-05 7.735741e-07
[11] 3.164621e-08 7.846168e-10 8.916100e-12
> round(dbinom(x,12,0.12),digits=4)
 [1] 0.2157 0.3529 0.2647 0.1203 0.0369 0.0081 0.0013 0.0001 0.0000 0.0000
[11] 0.0000 0.0000 0.0000
```

The most probable no. of faulty items is 1.

(o) Solve by R: how many faulty items can we assure with a probability of at least 0.9 from the seller's point of view?

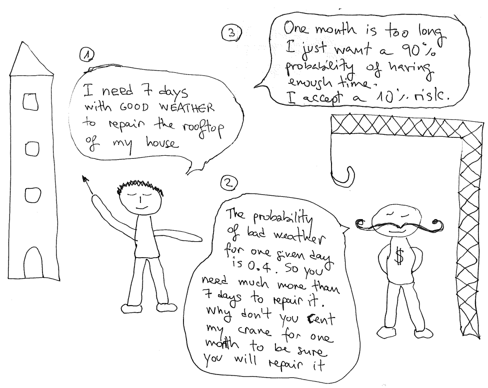
Assuring a no. of faulty items from the seller's pov is giving a maximum no. of faulty items with a given probability:

$$P[X \leq x] > 0.9$$

```
> x=0:12
> pbinom(x,12,0.12)
 [1] 0.2156712 0.5685876 0.8332749 0.9535873 0.9905014 0.9985554 0.9998367
 [8] 0.9999864 0.9999992 1.0000000 1.0000000 1.0000000 1.0000000
```

The first x value reaching a 0.9 probability is $x = 3$. So the seller can assure with a 0.9 probability that the numer of faulty items will be 3 or lesser.

[76] In a given place the probability of raining is 0.4 and it's assumed total independence among different days. In order to construct a rooftop we need 7 days without rain. For how many days should we rent a crane to construct the rooftop with a 0.9 probability? And with 0.99 probability?



Method: trial-and-error tuning

- **We start with renting for 7 days (X: raining days** $\sim B(n = 7, p = 0.4)$**):**

  P[**repair**] $= P$[**0 raining days among 7 days**] $= 0.6^7 = 0.4^0 0.6^7 \dfrac{7!}{0!7!} = 0.028$

  - **With R:**

    ```
    >dbinom(0,7,0.4)
    [1] 0.0279936
    ```

  Probability of repairing is too small. So we try with renting for one more day, in total 8 days (X: raining days $\sim B(n = 8, p = 0.4)$**):**

  P[**repair**] $= P$[**at most 1 raining day among 8 days**] $= P[X \leq 1] = P[X = 0] + P[X = 1] = 0.4^0 0.6^8 \dfrac{8!}{0!8!} + 0.4^1 0.6^7 \dfrac{8!}{1!7!} = 0.106$

  - **Using R:**

    ```
    > pbinom(1,8,0.4)
    [1] 0.1063757
    ```

  We don't reach 0.9.

- **We try with one more day: 9 days (X: raining days** $\sim B(n = 9, p = 0.4)$**):**

  P[**repair**] $= P$[**at most 2 raining day among 9 days**] $= P[X \leq 2] = P[X = 0] + P[X = 1] + P[X = 2] = 0.23$

  - **Using R:**

    ```
    > pbinom(2,9,0.4)
    [1] 0.231787
    ```

- **And so on, until we reach at least a 0.9 probability. Using R:**

  ```
  > x=0:15;y=7:22
  > pbinom(x,y,0.4)
   [1] 0.0279936 0.1063757 0.2317870 0.3822806 0.5327742 0.6652086 0.7711560
   [8] 0.8498599 0.9049526 0.9416811 0.9651873 0.9797184 0.9884371 0.9935341
  [15] 0.9964467 0.9980778
  ```

  We reach a 0.9 probability with 15 days.

  For a 0.99 probability of repairing the rooftop, we need logically more days: 20 days, exactly.

- **A complementary question. How many days do we need on average to repair the rooftop?**

[77] We sell batches of 10 pieces. One of our customers inspects all the pieces and if he finds one faulty piece or more in one batch he rejects it. The probability of a piece being faulty is 0.15. On the other side, the customers buys 25 sets of batches along the year and if he rejects more than 5% of then, he will cancel the contract.

i. What is the probability of the contract being cancelled?

ii. Which should the probability of a piece being faulty in order to be the probability of rejecting one batch at most 0.3?

(a)

X**:faulty items** $\sim B(n = 10, p = 0.15)$**.**

$$P[reject\ batch] = P[X \geq 1] = 1 - P[X = 0] = 1 - 0.15^0 0.85^{10} \frac{10!}{0!10!} = 0.804$$

Y**:rejected batches** $\sim B(n = 25, p = 0.804)$**:**

$$P[cancel\ contract] = P[Y > 25 \times 0.2] = P[Y > 5] = 1 - P[Y \leq 5] \approx 1$$

(b)

$$P[batch\ reject] = P[X \geq 1] = 1 - P[X = 0] = 1 - p^0 (1-p)^{10} \frac{10!}{0!10!} = 0.3$$

$$(1-p)^{10} = 0.7 \rightarrow 10 \ln(1-p) = \ln 0.7 = -0.35 \rightarrow \ln(1-p) = -0.035 \rightarrow 1 - p = exp(-0.035) \rightarrow p = 0.035$$

[78] In a flight 25% of tickets are cancelled or become vacant. For a given flight we have 12 seats. How many tickets should we sell in order to be the probability of having overbooking at most 0.15?

Let's remember that overbooking happens when a company books more tickets than seats in the airplane, as they hope that some of the travelers that booked the flight at the end will cancel it, but at the end there are more travelers at the airport than seats in the airplane, so some travelers are thrown off. The goal of this politics is to maximize the revenue, but overbooking remains a problem for the company and they will alway want to limit its probability.

[84] Last year there were 15 accidents in a road and 10 of them happened on Sundays and bank holidays. Should we conclude accidents become more frequent on those days? Remark: there were 82 Sundays and other holidays last year. Significance level: 1%.

The probability of a accident happening on a holiday when all days are the same regarding the no. of accidents is $p = \frac{82}{365}$**.**

They are asking if accidents are more frequent on holidays ($p > \frac{82}{365}$)**, so we take the opposite as the null hypothesis:**

$$H_0 : \textbf{on holidays accidents are more frequent} : p \leq \frac{82}{365}$$

The evidence also supports that accidents are more frequent on holidays. So, with this criterium too, in order to be cautious, we take the opposite as the null hypothesis or starting point:

$$H_0 : \textbf{on holidays accidents are more frequent} : p \leq \frac{82}{365}$$

Now, under the null hypothesis, that is, taken $p = \frac{82}{365}$**, we calculate the probability of the evidence or something stranger. We see in data that accidents are** *more frequent* **on holidays** ($10 > 15 \times \frac{82}{365}$)**; so the direction of the test will be on the upper side.**

Accidents on holidays distribute in this manner: $X \sim B\left(n = 15, p = \frac{82}{365}\right)$**, under** H_0**. So,**

$$P[X \geq 10] = 0.0003 < \alpha \rightarrow reject\ H_0$$

With R software:

```
> 1-pbinom(9,15,82/365)
[1] 0.0003156147
```

So, we have strong enough evidence to state that accidents are more frequent on holidays.

[85] There are 4 workers in a factory and each of them produces 6 pieces a day. Among the 24 pieces produced in a day we found 6 faulty pieces and only one of them was produced by Peter, the eldest worker. Should we decide he produces better that his colleagues? Significance level: 1%.

The probability of a piece being produced by Peter when all workers have the same experience or skills is 1/4=0.25.

They are asking is Peter does better ($p < 0.25$), so we take the opposite as the null hypothesis:

H_0: the workers are the same: $\mathbf{p} \geq \frac{1}{4} = \mathbf{0.25}$

According to other criteria too we should take the same null hypothesis: evidence shows that probability of a faulty piece being produced by the eldest worker is $\frac{1}{6} = 0.16$; so, it supports the second hypothesis. So, with caution, we take the first one as H_0 or the null hypothesis.

Now, we calculate the probability or evidence or something stronger. We perform the test because Peter produces a *small* number of faulty pieces compared to the other workers under the null hypothesis ($1 < 6 \times 0.25 = 1.5$), so the direction of the test (we also call it the critical region) is on the lower side.

The no. of faulty pieces produced by Peter distributes like this: $X \sim B(n = 6, p = 0.25)$,under H_0. So,

$$P[X \leq 1] = 0.53 > \alpha \rightarrow accept\ H_0$$

With R software:

```
> pbinom(1,6,0.25)
[1] 0.5339355
```

So, the evidence is not strong enough to state that Peter does better.

[86] A firm has designed a path to deliver packages. In order to program that in the most efficient way, the firm assumes that the probability for a package to be delivered at point A must be exactly 0.09. Among the last 14 packages, 4 of them were delivered at point A. Should we conclude that the program for the path should be remade? Significance level: 1%.

They are asking us if the program should be remade. So following both the 2nd criterium to give the null hypothesis, we set the null hypothesis as the opposite of the claim made in the question:

$$H_0 : p = 0.09$$

We reject the null hypothesis when the percentage of packages delivered at A point is very big or very small. Hence, it's a two-sided test.

But in this case the evidence shows that the proportion of delivered packages at A is 4/14=28%, so the data are pointing to the upper side:

$$P[evidence/H_0] = P[X \geq 4/p = 0.09] = \{X \sim B(14, 0.09)\} = 0.0314 > \alpha/2$$

We compare p-value with $\alpha/2$ because we are performing a two-sided test.

So we accept the null hypothesis and claim there is no reason to state that the program for the path should be remade.

[87a] We collected some daily sales in a shop:

46-87-56-64-55-67-64-65-72-75-98

(a) Test if population median is 60. Significance level: 10%. Calculate also the p-value

---

Critical value method

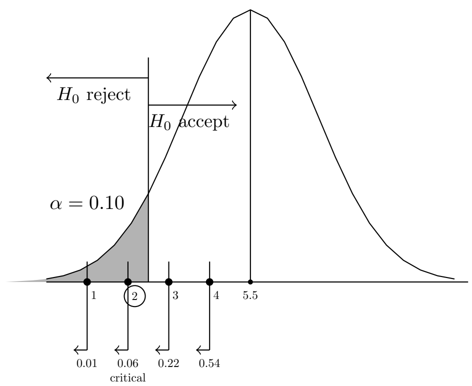
Signs: -+-+-+-++++++

$r^- = 3;\ r^+ = 8 \to r = 3$

Comment (not needed for solution): when the null hypothesis is absolutely true, $r = 11/2 = 5.5$.

Critical value ($n = 11, \alpha = 0.10$): $r^* = 2$

$\boxed{r = 3 > r^* = 2 \to H_0 : Me = 60\ accept}$



p-value method

$$
\begin{aligned}
p = P[r \leq 3] &= P[r=0] + P[r=1] + P[r=2] + P[r=3] \\
&= P[< Me, 0; > Me, 11] + P[< Me, 1; > Me, 10] + P[< Me, 2; > Me, 9] + P[< Me, 3; > Me, 8] \text{(or inverse!)} \\
&= \left(0.5^0 \cdot 0.5^{11} \cdot \frac{11!}{0!11!} + 0.5^1 \cdot 0.5^{10} + 0.5^3 \cdot 0.5^8 \ cdot \frac{11!}{3!8!} \right) \times 2 = 0.22
\end{aligned}
$$

$\boxed{p > \alpha \to H_0 : Me = 60\ accept}$

[89d] We have collected these test scores among a group of students:

67-78-75-81-82-77-72-69-87-75

Set 90% and 80% confidence intervals for the median.

Sample size: 10.

So we calculate cumulated probabilities for $B(n = 10, p = 0.5)$.

$P[X = 0] = 0.5^0 \cdot 0.5^{10} \cdot \frac{10!}{0!10!}$
$P[X = 1] = 0.5^1 \cdot 0.5^9 \cdot \frac{10!}{1!9!}$
$P[X = 2] = 0.5^2 \cdot 0.5^8 \cdot \frac{10!}{2!8!}$
....
$P[X = 10] = 0.5^{10} \cdot 0.5^0 \cdot \frac{10!}{10!0!}$

We can give the results using R:

```
> x=0:10
> round(dbinom(x,10,0.5),digits=4)
 [1] 0.0010 0.0098 0.0439 0.1172 0.2051 0.2461 0.2051 0.1172 0.0439 0.0098
[11] 0.0010
```

And now we cumulate them:

```
> cumsum(round(dbinom(x,10,0.5),digits=4))
 [1] 0.0010 0.0108 0.0547 0.1719 0.3770 0.6231 0.8282 0.9454 0.9893 0.9991
[11] 1.0001
```

We can calculate previous cumulated probabilities directly:

```
> round(pbinom(x,10,0.5),digits=4)
 [1] 0.0010 0.0107 0.0547 0.1719 0.3770 0.6230 0.8281 0.9453 0.9893 0.9990
[11] 1.0000
```

We order data: 67, 69, 72, 75, 75, 77, 78, 81, 82, 87

For a confidence interval of 90%, we leave 5% on each tail or extreme. So, we take the value that gives a maximum lower probability of 5% and the same for the upper side: the lower value leaving below it a maximum of 5% is $k = 1$. So $k + 1 = 2$, and we take the 2nd smallest and largest values: 69 and 82.

So we can assure with a $(1 - 2 \times 0.0107) = 98\%$ confidence level that the median in the population is between 69 and 82.

Let's give now the 80% confidence interval. In this case, we leave 10% on each tail, so we take $k = 2$ and then $k + 1 = 3$. So the interval for the median is 72-81 with $(1 - 2 \times 0.0547) = 89\%$ confidence level.

[90] The probability of a faulty piece being produced is 0.12. (a) Which is the distribution of the number of produced faultless pieces before the first produced faulty piece?

We are counting the number of faultless before the first faulty. Hence, p parameter is the probability of a piece being faulty: $p = 0.12$. The X number of faultless before the first faulty distributes in this way:

$$X \sim G(p = 0.12) : P[X = x] = 0.88^x 0.12; x = 0, 1, 2, 3, ...$$

(b) What is the probability of having 6 faultless pieces just before the first faulty piece?

Applying the previous formula:

$$P[X = 6] = 0.88^6 0.12$$

Or thinking about it:

$$P[X = 6] = P[000000X] = 0.88 \times \cdots \times 0.88 \times 0.12 = 0.88^6 0.12$$

(c) What is the probability of having 2 faultless pieces or less just before the first faulty piece?

Applying the previous formula

$$P[X \leq 2] = P[X = 0] + P[X = 1] + P[X = 2] = 0.88^0 0.12 + 0.88^1 0.12 + 0.88^2 0.12$$

(d) What is the expected number of faultless pieces before the first faulty piece?

As we have seen in the previous questions the number of faultless pieces till the first faulty one may take different values and so it must have an average value:

$X \sim G(p = 0.12) : E[X] = \frac{0.88}{0.12} = 7.33$ faultless pieces are expected till the first faulty piece.

(e) What is the probability of having 2 faulty pieces before the first faultless piece?

Now we wait for the first faultless, so $p = 0.88$, the probability for a faultless piece:

$$X \sim G(0.88) : P[X = 2] = 0.12^2 0.88$$

(f) What is the expected number of faulty pieces before the first faultless piece?

$X \sim G(0.88) : E[X] = \frac{0.12}{0.88} = 0.13$ faulty items are expected till the first faultless one.

[92]

i. **Calculate the probability of having 3 faulty pieces among those 5 pieces, both by means of the binomial coefficientes and multiplying simple probabilities.**

ii. **Calculate the mean number of faulty pieces.**

iii. **Calculate the probabilities of all possible values for the number of faulty pieces. Calculate the mode by both examining the previous results and applying the formula for the mode.**

iv. **State the probabilities equal to the probability of having 3 faulty pieces (that is to say, the symmetries)**

v. **Calculate the variance of the number of faulty pieces, and compare it to the variance of faulty pieces, when pieces are drawn with devolution.**

vi. **Calculate the probability of the faulty pieces being 60%, both when we draw 5 pieces and 25 pieces. Do we have a paradox with these results?**

```
> x=0:5
> round(dhyper(x,20,80,5),digits=4)
[1] 0.3193 0.4201 0.2073 0.0478 0.0051 0.0002
```

So mode is $x = 1$ faulty pieces, that is, the value with the biggest probability.

Some people could think both probabilities should be the same, as we are calculating probabilities about the same percentage. But that's not true.

- $P[\text{of 5 pieces \%60 faulty}] = P[X = 5 \times 0.6 = 3] = 0.0478$, as it was given in the previous section.
- $P[\text{of 25 pieces \%60 faulty}] = P[X = 25 \times 0.6 = 15] = 0.000000006428355$.

Both probabilities are quite different. Let's remark the 60% percentage is far away from the 20% inside the box. So we can say 60% is something strange. And strange things occur more probably in the short term than in the long term: so it's more probable to have 60% with 5 pieces drawn than with 25 pieces drawn (remark: because of the same reason, it's more probable to get 66 with a dice than 666666).

Remark: we calculate the second probability with R software:

```
> dhyper(15,20,80,25)
[1] 1.052585e-07
```

[97] On average 6 customers per minute enter a shop, randomly and with independence from one to another. Each customer needs 2 minutes to pay. How many cashiers should we hire in order to have a 0.9 probability of not having a queue?

If there are 9 cashiers, the queue is created when 10 customers or more have arrived in 2 minutes. So, for not having a queue, number of customers must be equal or less than the number of cashiers. If we take X number of customers in 2 minutes eta x the number of cashiers:

$$P[\text{not having a queue}] = P[X \leq x] = 0.9$$

For two minutes, $\lambda = 12$, and so we take the first x value giving a probability bigger than 0.9. Intuitively, if on average we have 12 customers in 2 minutes, x will be bigger than 12. So, we begin with $x = 13$:

```
>ppois(13:20,12)
[1] 0.6815356 0.7720245 0.8444157 0.8987090 0.9370337 0.9625835 0.9787202 [8] 0.9884023
```

Hence, the number of cashiers must be at least 17 to have a probability of at least 0.9 of not having a queue in 2 minutes.

[98] We are managing industrial process, the probability of producing a faulty item is 0.0022. We have sold a batch of 4000 items, but we'll return it whenever we find more than 6 faulty items.

(a) What is the probability of returning the batch? Calculate both by means of the binomial distribution and the Poisson distribution.

(b) Among 1000 items, how many faulty items can we ensure with a 0.9 probability from the seller's point of view?

(c) How many faulty items can be assured with a 0.9 probability from the customer's point of view ?

(a)

The number of faulty items (X) in the batch distributes in this way:

$$X \sim B(n = 4000, p = 0.0022) \to P(\lambda = np = 8.8)$$

$\lambda = 8.8$ means that we have 8.8 faulties on average among 4000 items.

$$P[return\ batch] = P[X > 6]$$

By means of the binomial distribution:

$$\begin{aligned} P[X > 6] = P[X = 7] + P[X = 8] + \cdots + P[X = 4000] = \\ &= 0.0022^7 0.9978^{3993} \frac{4000!}{7!3993!} + 0.0022^8 0.9978^{3992} \frac{4000!}{8!3992!} + \cdots + 0.0022^0 0.9978^{4000} \frac{4000!}{0!4000!} = \\ &= 0.7746894 \end{aligned}$$

By means of the Poisson distribution:

$$\begin{aligned} P[X > 6] &= 1 - P[X \le 6] = 1 - (P[X = 0] + P[X = 1] + \cdots + P[X = 6]) = \\ &= 1 - \left(\frac{e^{-8.8} 8.8^0}{0!} + \frac{e^{-8.8} 8.8^1}{1!} + \cdots + \frac{e^{-8.8} 8.8^6}{6!} \right) = \\ &= 0.7743897 \end{aligned}$$

As we see, the Poisson approximation is quite good, with a error smaller than 1/1000.

R commands to calculate those results are:

```
> pbinom(6,4000,0.0022,lower.tail=F)
[1] 0.7746894
> 1-ppois(6,8.8)
[1] 0.7743897
```

[99] The probability of a worker having an accident in a year is 0.0012.

(a) On average, how many accidents happen among 1000 workers?

(b) In a factory there are 3000 workers and there were 7 accidents last year. Should we conclude that prevention is not as successful as in the other factories? Significance level: 0.05.

The number of workers having an accident among 1000 workers (X) distributes in this way:

$$X \sim B(n = 1000, p = 0.0012) \rightarrow P(\lambda = np = 1.2)$$

(a)

The number of workers having an accident is on average 1.2 per year, given by λ.

(b)

H_0: prevention or safety is normal: $\lambda \leq 1.2 \times 3 = 3.6$

In order to set the null hypothesis, the criterion is to take the opposite of that they ask or suspect in the question.

To calculate the corresponding p value:

$$P[X \geq 7/\lambda = 3.6] = 0.07327342 > \alpha$$

This result is calculated in R by this command:

```
> ppois(6,3.6,lower.tail=F)
[1] 0.07327342
```

Hence, we accept the null hypothesis and decide that there is not strong enough evidence to support that prevention is not as successful as in the other factories.

Remark: the test is one-sided as the statement given in the question is given only in one direction.

[101] A bridge has been finished this year. The contractors think that a flood that would wreck the bridge should happen only once in 1000 years.

i. What is the probability of the bridge standing in the next 100 years?

ii. For how many years will the bridge stand with a 0.95 probability?

(a)

Fron the definition of return period: $\lambda_{1000 \text{ urte}} = 1$

Now we have to adjust the lambda parameter to 100 years, as the asked probability refers to that period: $\lambda_{100 \text{ urte}} = 0.1$

The bridge standing means that there will be no flood. So:

$$P[X=0] = \frac{e^{-0.1} 0.1^0}{0!} = 0.9048$$

(b)

Comparing to the previous result, we can say that the bridge will stand with a probability of 95% for less than 100 years. Let's calculate the exact period for that probability:

$$P[X=0] = \frac{e^{-\lambda}\lambda^0}{0!} = e^{-\lambda} = 0.95$$
$$\lambda = -\ln 0.95 = 0.05129$$

By cross-multiplication:
$100 \rightarrow \lambda = 0.1$
$t \quad \rightarrow \lambda = 0.05129$

Then, $t = 51.29\ years$.

[106] Customers come randomly and independently to a queue, with a mean time till the next customer of 5 minutes. It turns out that the next customer has come in 20 minutes. Should we reject the 5 minutes average? Significance level: %1.

Customers coming randomly and independently means that number of customers is Poisson distributed, and time between customers exponential distributed.

Mean time is $\frac{1}{\lambda} = 5min \to \lambda = \frac{1}{5}$ *per min.*

By caution, we take the null hypothesis as nothing has changed, as all is going as usual:

$$H_0 : \frac{1}{\lambda} = 5min$$

Evidence tells us that next customer came in 20 minutes. Given H_0, that's very much; we are amazed because 20 minutes is a big number. So, we reject H_0 on the *upper side*:

$$P[T > 20min] = e^{-\frac{1}{5}\times 20} = 0.018 > \alpha \to accept\ H_0$$

There is no strong enough evidence to support that the average time to the next customer is bigger than 5 minutes.

[108] The number of daily sales of washing machines at a a retailer's is between 0 and 9, following a discrete uniform distribution. Calculate:

(a) The probability of selling on a given day 3 machines or less.

(b) The mean number and variance of machines sold.

(c) The probability of a given sequence of sales along 4 days (e.g., 2,6,1,8). Give the number of days, the minimum and the maximum for that sequence.

(d) Along 4, 5 and 6 days respectively, the probability of the maximum being exactly 8, as in the given sequence.

(e) The shop assistant is able to manage only 7 machines one day. Give the probability of not being able to manage all the sales at least one day along the 4 days.

(f) Along 4, 5 and 6 days respectively, the probability of the minimum being 1, as in the given sequence.

(g) There are losses when the number of daily sales is smaller than (or equal to) 2. Give the probability of having at least one day along those 4 days of being losses.

(h) How many days do we need in order for the probability of the maximum being 9 to be at least 0.95? And in order for the probability of the maximum being 9 or less to be at least 0.95?

(i) How many days do we need in order for the probability of the maximum being 8 to be at least 0.95? And in order for the probability of the maximum being 7 with the same probability?

(j) Which should be the top value of the distribution, instead of 9, in order for the maximum along 4 days to be at least 15 with a 0.95 probability? Idem, with a 0.8 probability. And along 8 days with both probabilities?

(a)

$$P[X = x_i] = \frac{1}{10}\ ;\ x_i = 0, 1, \ldots, 9$$

So,

$$P[X \leq 3] = P[X = 0 + P[X = 1] + P[X = 2] + P[X = 3] = \frac{1}{10} + \frac{1}{10} + \frac{1}{10} + \frac{1}{10} = 0.4$$

Or, another way: $P[X \leq 3] = \frac{i}{n} = \frac{4}{10}$

(b)

$$X \sim U(0, 1, \ldots, 9) \rightarrow \mu = \frac{0+9}{2} = 4.5 \text{and} \sigma^2 = \frac{(9-0+2)(9-0)}{12} = \frac{99}{12}$$

The expected value is exactly on the middle way, as all the possible values have the same probability.

(c)

The values in a given sequence are linked by AND (2 and 6 and 1 and 8), so we multiply the simple probabilities.

$$P[2, 6, 1, 8] = \frac{1}{10} \times \frac{1}{10} \times \frac{1}{10} \times \frac{1}{10} = \frac{1}{10000}$$

$n = 4$, $X_{max} = 8$, $X_{min} = 2$

(d)

n=4 days

$8 \rightarrow i = 9$ (8 is the 9th value)

$$P[X_{max} = 8] = \left(\frac{i}{N}\right)^n - \left(\frac{i-1}{N}\right)^n = \left(\frac{9}{10}\right)^4 - \left(\frac{8}{10}\right)^4 = 0.2465$$

n=5 days

$$P[X_{max} = 8] = \left(\frac{9}{10}\right)^5 - \left(\frac{8}{10}\right)^5 = 0.26281$$

n=6 days

$$P[X_{max} = 8] = \left(\frac{9}{10}\right)^6 - \left(\frac{8}{10}\right)^6 = 0.269297$$

The probability of getting a maximum of 8 (close to the absolute maximum, that is, to 9) is bigger as the number of days increases.

(e)

The probability of not being able to do the work in one day is 0.2 (when sales are 8 or 9). Hence, the number of days X in which the assistant is not able to do the work distributes $B(n = 4, p = 0.2)$. So:

$P[\text{at least one day not able}] = P[X = 1] + P[X = 2] + P[X = 3] + P[X = 4] = 1 - P[X = 0] = 1 - 0.8^4 = 0.5904$

Taking the long method, we should calculate in this way. For example: $P[X = 1] = 0.2^1 \cdot 0.8^4 \cdot \frac{4!}{3!1!}$

Anoter way of doing it: he/she is not able to do work at least one day when maximum is 8 or more. By means of the formula:

$$P[\text{not able to complete the work}] = P[X_{max} \geq 8] = 1 - P[X_{max} \leq 7] = 1 - \left(\frac{8}{10}\right)^4 = 0.5904$$

(f)

1 is the 2nd value, so $i = 2$.

n=4 days

$$P[X_{min} = 1] = \left(\frac{N-(i-1)}{N}\right)^n - \left(\frac{N-i}{N}\right)^n = \left(\frac{10-(2-1)}{10}\right)^4 - \left(\frac{10-2}{10}\right)^4 = 0.2465$$

n=5 days

$$P[X_{min} = 1] = \left(\frac{10-(2-1)}{10}\right)^5 - \left(\frac{10-2}{10}\right)^5 = 0.26281$$

n=6 days

$$P[X_{min} = 1] = \left(\frac{10-(2-1)}{10}\right)^6 - \left(\frac{10-2}{10}\right)^6 = 0.269297$$

The probability of getting a minimum of 1 (close to the absolute maximum, that is, to 0) is bigger as the number of days increases. We should also remark that the probabilities are analogous to those of the maximum being 8. That's because 1 and 8 are symmetric (second values from the top and from the beginning).

(g)

The probability of having losses in one day is 0.3 (when sales are 0,1 or 2). Hence, the number of days X in which we have losses distributes $B(n=4, p=0.3)$. So:

$P[\text{at least one day losses}] = P[X=1]+P[X=2]+P[X=3]+P[X=4] = 1-P[X=0] = 1-0.7^4 = 0.7599$

Taking the long method, we should calculate in this way. For example: $P[X=1] = 0.3^1 \cdot 0.7^4 \cdot \frac{4!}{3!1!}$

Another way to solve it: we have losses on at least one day when the minimum along 4 days is 2 or less. Using the correpsonding formula:

$$P[X_{min} \leq 2] = 1 - P[X_{min} \geq 3] = 1 - \left(\frac{N-(i-1)}{N}\right)^n = 1 - \left(\frac{10-(4-1)}{10}\right)^4 = 0.7599$$

(h)

- $9 \rightarrow i = 10$:

$$P[X_{max} = 9] = \left(\frac{i}{N}\right)^n - \left(\frac{i-1}{N}\right)^n = \left(\frac{10}{10}\right)^n - \left(\frac{9}{10}\right)^n = 1 - 0.9^n = 0.95$$

$$1 - 0.9^n = 0.95 \rightarrow 0.9^n = 0.05 \rightarrow n \ln 0.9 = 0.05 \rightarrow n = \frac{\ln 0.05}{\ln 0.9} = 28.43 \rightarrow 29 \; days$$

- It's sure for any number of days, with 1 probability, that maximum will be 9 or less.

(i)

- $8 \rightarrow i = 9$:

$$P[X_{max} = 8] = \left(\frac{i}{N}\right)^n - \left(\frac{i-1}{N}\right)^n = \left(\frac{9}{10}\right)^n - \left(\frac{8}{10}\right)^n = 0.95$$

We cannot solve with logarithms: $\ln(a-b) \neq \ln a - \ln b$. So, trying for different values with R software:

```
> n=seq(1,20,by=1)
> 0.9^n-0.8^n
 [1] 0.1000000 0.1700000 0.2170000 0.2465000 0.2628100 0.2692970 0.2685817
 [8] 0.2626951 0.2532028 0.2413043 0.2279113 0.2137101 0.1992110 0.1847875
[15] 0.1707068 0.1571545 0.1442538 0.1320802 0.1206737 0.1100474
```

Cannot reach 0.95. Maximum probability at $n = 7$.

- $P[X_{max} \geq 8] = P[X_{max} = 8] + P[X_{max} = 9] = \left[\left(\frac{9}{10}\right)^n - \left(\frac{8}{10}\right)^n\right] + \left[\left(\frac{10}{10}\right)^n - \left(\frac{9}{10}\right)^n\right] = 0.95$

With R:

```
> n=seq(1,20,by=1)
> (0.9^n-0.8^n)+(1-0.9^n)
 [1] 0.2000000 0.3600000 0.4880000 0.5904000 0.6723200 0.7378560 0.7902848
 [8] 0.8322278 0.8657823 0.8926258 0.9141007 0.9312805 0.9450244 0.9560195
[15] 0.9648156 0.9718525 0.9774820 0.9819856 0.9855885 0.9884708
```

So, solution is $n = 14$.

Conclusion: it's not always possible to reach a given probability for an exact value for the maximum (or minimum) but it's always possible to get any probability (as big as we want) for a minimum value for the maximum (or a maximum value for the minimum).

(j)

n=4,p=0.95

We name the top value of the uniform distribution x_N. As the first value is 0, $N = x_N + 1$ (for example, 9 is the 10th value). So let's calculate firstly N, which is the parameter we use in formulas:

$$P[X_{max} \leq x_i] = \left(\frac{i}{N}\right)^n$$

$$P[X_{max} \geq 15] = 1 - P[X_{max} \leq 14] = 1 - \left(\frac{15}{N}\right)^4 = 0.95$$

$$1 - \left(\frac{15}{N}\right)^4 = 0.95 \rightarrow \ln N = \ln 15 - \frac{\ln 0.05}{4} = 3.45 \rightarrow N = 31.5 \rightarrow 32 \rightarrow x_N = 31$$

n=4,p=0.8: $\ln N = \ln 15 - \frac{\ln 0.2}{4} = 3.11 \rightarrow N = 22.42 \rightarrow 23 \rightarrow x_N = 22$

n=8,p=0.95: $\ln N = \ln 15 - \frac{\ln 0.05}{8} = 3.08 \rightarrow N = 21.75 \rightarrow 22 \rightarrow x_N = 21$

n=8,p=0.8: $\ln N = \ln 15 - \frac{\ln 0.2}{8} = 2.90 \rightarrow N = 18.17 \rightarrow 19 \rightarrow x_N = 18$

It's quite logical: with bigger x_N top values (bigger dice), it's more probable to reach a given value for the maximum. And along more days (more trials with the dice) it's also more probable to get a big value for a maximum (or more).

[110] We only know the price growth of an item will be 0 to 10 in the next year.
(a) What is the probability of the price growth being bigger than 6 (give the answer both by reasoning and calculation).
(b) Which is the expected price growth? Give the answer both by reasoning and calculation.
(c) Which is the variance of the price growth?

As we *only* know the interval of sales, we have absolute uncertainty about sales and we should give all them the same probability. So, continuously: $X \sim U(a = 0, b = 10)$.

(a)

Applying the distribution function:

$F(x) = P[X < x] = \frac{x - a}{b - a} = \frac{x}{10}$; $0 < x < 10$.

$$P[X > 6] = 1 - P[X < 6] = 1 - \left[\frac{6}{10}\right] = 0.4$$

Reasoning: as all the values have the same probability, the probability of sales being larger than 6, that is being 6 to 10, will be proportional to the distances between those values: from 6 to 10 we have 4, and total distance is 10-0=10. So, probability will be 4/10=0.4.

(b)

$$\mu = \frac{a + b}{2} = \frac{0 + 10}{2} = 5$$

Logical: exactly on the middle way from 0 to 10, as all those values have the same probability.

(c)

$$\sigma^2 = \frac{(b - a)^2}{12} = \frac{(10 - 0)^2}{12} = 8.33$$

---

[111] $X \sim U(0, 8)$**. Calculate** fz **so that** $P[X > \mu_X + z] = 0.1$**.**

$$\mu_X = \frac{0 + 8}{2} = 4$$

$$f(x) = \frac{1}{8}$$

$$P[X > 4 + z] = 0.1 \rightarrow \int_{4+z}^{8} \frac{1}{8} dx = \left[\frac{x}{8}\right]_{4-z}^{8} = \left[\frac{8}{8}\right] - \left[\frac{4 + z}{8}\right] = 0.1 \rightarrow \frac{8 - (4 + z)}{8} = 0.1 \rightarrow z = 3.2$$

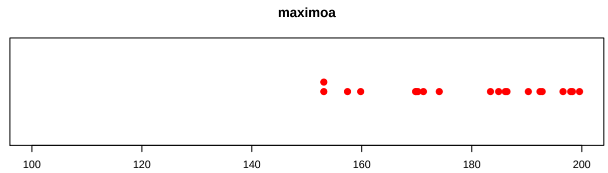
[114] Sales in a shop distribute between 100 and 200, and furthermore we have absolute uncertainty about the sales number in that interval.

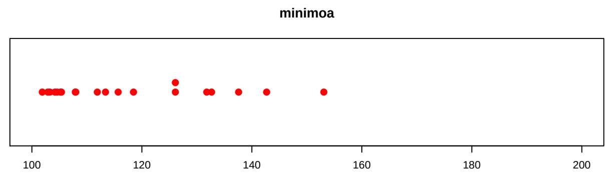
- **i. Simulate sales numbers along 4 days.**
- **ii. Give the maximum, minimum and range for each 4 days sequence simulated previously.**
- **iii. Repeat (a) and (b) 20 times.**
- **iv. Depict maximum, minimum and range values in a suitable chart.**
- **v. Interpret the charts: which are the mean values, appproximately, for maximum, minimum and range along 4 days?**
- **vi. Calculate exact expected values for maximum, minimum and range along 4 days.**
- **vii. Explain the previous expected values in a intuitive way.**
- **viii. Predict and calculate the probability of the maximum being smaller than 140.**
- **ix. Predict and calculate the probability of the minimum being larger than 180.**
- **x. Predict and calculate the probability of the range being smaller than 10.**

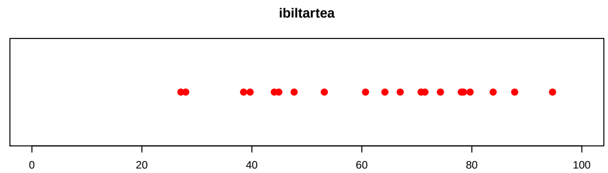
(a), (b), (c) eta (d)

With R software:

```
> a=matrix(round(runif(20*4,100,200),digits=1),ncol=4)
> a
       [,1]  [,2]  [,3]  [,4]
 [1,] 187.5 198.0 196.0 153.1
 [2,] 153.1 105.4 115.5 122.8
 [3,] 170.2 126.1 131.4 163.3
 [4,] 137.6 152.1 186.1 198.3
 [5,] 161.5 162.6 132.7 171.2
 [6,] 108.0 164.8 186.1 148.7
 [7,] 150.4 182.7 107.9 186.4
 [8,] 152.7 190.3 126.1 166.0
 [9,] 105.2 184.9 168.1 168.7
[10,] 183.4 132.4 176.7 111.9
[11,] 192.4 176.4 186.1 104.6
[12,] 131.8 141.7 159.8 134.5
[13,] 192.8 165.8 118.5 146.7
[14,] 139.8 184.9 101.9 196.6
[15,] 169.5 169.8 149.4 142.7
[16,] 115.7 129.7 140.1 199.6
[17,] 102.9 169.9 165.1 108.9
[18,] 104.2 157.4 147.9 135.6
[19,] 174.1 114.8 119.1 103.3
[20,] 153.1 137.4 127.1 113.4
> max=do.call(pmax, as.data.frame(a))
> max
 [1] 198.0 153.1 170.2 198.3 171.2 186.1 186.4 190.3 184.9 183.4 192.4 159.8
[13] 192.8 196.6 169.8 199.6 169.9 157.4 174.1 153.1
> min=do.call(pmin, as.data.frame(a))
> min
 [1] 153.1 105.4 126.1 137.6 132.7 108.0 107.9 126.1 105.2 111.9 104.6 131.8
[13] 118.5 101.9 142.7 115.7 102.9 104.2 103.3 113.4
> ibil=max-min #range
> ibil
 [1] 44.9 47.7 44.1 60.7 38.5 78.1 78.5 64.2 79.7 71.5 87.8 28.0 74.3 94.7 27.1
[16] 83.9 67.0 53.2 70.8 39.7
```







(e)

Average maximum along 4 days is around 180, average minimum around 120, and average range around 10.

(f)

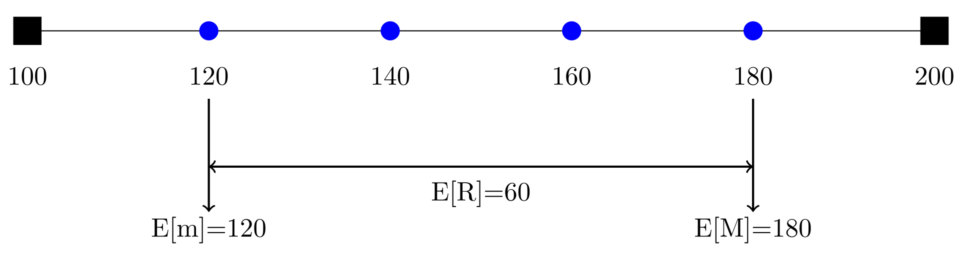
$$E[M] = 100 + \frac{4 \times (200 - 100)}{4 + 1} = 180$$

$$E[m] = 100 + \frac{(200 - 100)}{4 + 1} = 120$$

$$E[R] = E[M] - E[m] = 60$$

(g)

Previous results are intuitive: we expect along 4 days to be big, mid, and small sales. Ideally, sales should be 120, 140, 160 and 180, as probability is uniform or equal along the interval. So we expect the maximum to be 180, them minimum 120, and the range 10.



(h) As we see in the simulations there's no maximum value below 140, so the probability must be very small. In fact, if we want the maximum to be smaller than 140, all sales must be smaller than 140:

$$P[M < 140] = 0.4^4 = 0.0256$$

(i) As we see in the simulations there's no minimum value above 180, so the probability must be very small. In fact, if we want the minimum to be larger than 180, all sales must be larger than 180:

$$P[m > 180] = 0.2^4 = 0.0016$$

(i) As we see in the simulations there's no range value below 10, so the probability must be very small. In fact, if we want the range to be smaller than 10, all sales must be in a interval of width 10 or shorter:

$$P[R < 10] = 4 \times \left(\frac{10}{100}\right)^3 \times \frac{96}{100} + \left(\frac{10}{100}\right)^4 = 0.00394$$

---

[115] Sales in a shop distribute $U(0, b)$, b being an unknown parameter. To estimate b, we take the maximum of sales along n days. E.g., if we get 23-86-176-158, we estimate $\hat{b} = 176$. It's clear that the estimation, that is to say, the maximum will always be below b, so we have a systematic error. But we can control it.

i. **In percentage, how much error do we make taking the maximum sales number along 7 days as an estimation for b?**

ii. **Along how many days should we collect sales data if we want the error to be at most 5%?**

iii. **What should we do to fix that error?**

iv. **We assume that sales distribute $U(100, 200)$. Along 4 days the maximum was 140. Should we conclude that the maximum of the distribution is really 200?** $\alpha = 0.1$

(a)

Along 7 days this is the expected value for the maximum: $E[\hat{b}] = E[M] = 0 + \dfrac{7 \times (b - 0)}{8} = \dfrac{7b}{8}$

Respect to b parameter, this is the mean error: $e = b - \dfrac{7b}{8} = \dfrac{b}{8}$

As a percentage: $\dfrac{b/8}{b} = \dfrac{1}{8} = 0.125 = \%12.5$

Briefly, mean maximum along 7 days is 12.5% above from b.

(b)

Clearly, we will need data along more days.To have an error of 5% , absolute error must be $0.05b = b \times 1/20$ and so the mean maximum has to be $19b/20$. So, we will need data along 19 days.

(c)

To have a null mean error, we just have to multiply the maximum with the corresponding percentage.

As the mean maximum is $19b/20$, and the parameter to estimate is b, in order to reach from $19b/20$ on average, to b on average:

$$\frac{19b}{20} \times k = b \rightarrow k = \frac{20}{19} = 1.052$$

So, we have to multiply the sample maximum along 20 days by 1.052 in order the get *on average* the true value of the unknown b parameter. Let's remark that we will never know how much the exact error, we just know that with $1.052M$ we will reach on average the true value of b; but in fact, everytime we use $1.052M$ as an estimate for b, we may have an excess in the estimation (perhaps M is the true value por b, for example), give the exact value ($b = 1.052M$), or make an underestimation (when b is really bigger than $1.052M$). But we can only know if $1.052M$ is right as an estimate for b on average, that is to say, in the long term (after making many estimates), and absolutely not in each case.

(d)

$H_0 : U(100, b \geq 200)$ (it look like b is smaller than 200, so we take the opposite)

A 140 maximum is small, so we calculate the probability below, and compare it to α because it's a one-sided test (because we reject the null hypothesis anly when the maximum is small enough):

$$P[M < 140] = \left(\frac{140 - 100}{200 - 100}\right)^4 = 0.0256 < \alpha$$

So we reject $b = 200$ (or bigger) and decide that b must be smaller.

[117.] Tenperature into a freezer follows a $N(0,1)$ standard normal distribution. Calculate these probabilities both by means of the statistical table and R programming language commands:

(a) $P[Z < 1.42]$

(b) $P[Z < 3.98]$

(c) $P[Z < 5.6]$

(d) $P[Z > 2.75]$

(e) $P[Z < -0.68]$

(f) $P[Z > -1.02]$

(g) $P[0.48 < Z < 1.92]$

(h) $P[-1.24 < Z < -0.98]$

(i) $P[-2.19 < Z < 0.55]$

(j) Give the value in the standard normal distribution that accumulates below it a 95% probability.

(k) Give the value in the standard normal distribution that accumulates beyond it a 20% probability.

(l) Give the value in the standard normal distribution that accumulates below it a 10% probability.

With R software

```
> pnorm(1.42)
[1] 0.9221962
> pnorm(3.98)
[1] 0.9999655
> pnorm(5.6)
[1] 1
> pnorm(2.75,lower.tail=F) #Remark that we must take the exact value,
> #as the normal distribution is a continuous distribution
[1] 0.002979763
> pnorm(-0.68)
[1] 0.2482522
> pnorm(-1.02,lower.tail=F)
[1] 0.8461358
> pnorm(1.92)-pnorm(0.48)
[1] 0.2881847
> pnorm(-0.98)-pnorm(-1.24)
[1] 0.05605536
> pnorm(0.55)-pnorm(-2.19)
[1] 0.6945782
> qnorm(0.95) #we get the exact value (not an approximation),
> #as when we use the tables
[1] 1.644854
> qnorm(0.2,lower.tail=F)
[1] 0.8416212
> qnorm(0.1) #no need to think it's negative and so on,
>#as when we use the tables
[1] -1.281552
```

[121] Daily production in a factory follows the normal distribution, with a 8 ton average and a deviation of 1 ton. (a) What is the probability of producing less than 30 tons in 4 days? (b) How much production can we ensure with a 90% probability? (c) How many days do we need to fulfill a batch of 60 tons, if we want the probability of not fulfilling the deadline to be 15%? (d) Idem, with a 1% probability. (e) Taxes for producing will be 40.000 €, with a tax allowance of 1.000 per ton. What is the probability of paying for taxes more than 10.000 €?

(a) Daily production distributes as this:

$$X_i \sim N(\mu = 8, \sigma = 1)$$

Applying reproductivity of the normal distribution, production in 4 days distributes as this:

$$\mathbf{X} = X_1 + X_2 + X_3 + X_4 \sim N(\mu = 8 \times 4 = 32, \sigma = \sqrt{1^2 \times 4} = 2)$$

So, $P[\mathbf{X} < 30] = P\left[Z < \frac{30-32}{2}\right] = P[Z < -1] = P[Z > 1] = 1 - P[Z < 1] = 1 - 0.8413 = 0.1587.$

(b) Ensuring a production level is determining a production level *at least*:

$$P[X > x] = 0.9 \xrightarrow{standardizing} P\left[Z > \frac{x-32}{2}\right] = 0.9 \rightarrow \frac{x-32}{2} = -1.28 \rightarrow x = 29.44$$

We ensure that production in 4 days will be bigger than 29.44.

(c) Production in n days follows this distribution:

$$\mathbf{X} = X_1 + X_2 + \ldots + X_n \sim N(\mu = 8n, \sigma = \sqrt{1^2 \times n} = \sqrt{n})$$

For a deadline of n days, we don't fulfill it if the production for those n days is smaller than 60, and we want the probability of that being 0.15:

$$\begin{aligned} P[X < 60] = 0.15 \rightarrow (standardizing) P\left[Z < \frac{60-8n}{\sqrt{n}}\right] = 0.15 \rightarrow & \frac{60-8n}{\sqrt{n}} = -1.03 \\ & \rightarrow 8n - 1.03\sqrt{n} - 60 = 0 \\ & rightarrow(n = x^2) 8x^2 - 1.03n - 60 = 0 \\ & \rightarrow x = 2.80 \rightarrow n = x^2 = 7.84 \end{aligned}$$

As we have stated n as an integer number, we must take 7 or 8. With 7 days the probability of not fullfilling would be bigger than 15% , so we take 8 days in order to have a probability smaller than 15%.

(d) For a 0.01 probability for not fulfilling the deadline, we will need more days:

$$\begin{aligned} P[X < 60] = 0.15 \xrightarrow{standardizing} P\left[Z < \frac{60-8n}{\sqrt{n}}\right] = 0.01 \rightarrow & \frac{60-8n}{\sqrt{n}} = -2.32 \\ & \rightarrow 8n - 2.32\sqrt{n} - 60 = 0 \\ & \xrightarrow{n=x^2} 8x^2 - 2.32n - 60 = 0 \\ & \rightarrow x = 2.88 \rightarrow n = x^2 = 8.29 \end{aligned}$$

So we will need for that probability a deadline of 9 days.

(e) We calculate taxes in this way: $T = 40.000 - 1000X$.

Applying the property for linear transformations of the normal distribution, taxes distribute as this:

$$T \sim N(\mu = 40.000 - 1000 \times 8 = 32.000, \sigma = 1000 \times 1 = 1000)$$

Remark: adding or substracting constants don't change σ, multiplying constants multiply σ but by their absolute value.

$$P[T > 10.000] = P\left[Z > \frac{10.000 - 32.000}{1000}\right] = P[Z > -22] = P[Z < 22] \approx 1$$

[122] 12 people get into an elevator, and the maximum weight allowed is 900 kg. Which should the average weight of a person in order to not exceeding it being 0.9? Remark: standard deviation of the weight of a person is 10 kg.

Total weight for 12 people distributes in this way, applying the property of reproductivity of the normal distribution (W: total weight):

$$W = w_1 + w_2 + \cdots + w_n \sim N(12\mu, \sigma = \sqrt{12 \times 10^2} = 34.64)$$

To not exceed 900kg weight with a 0.9 probability:

$$P[W < 900] = P\left[Z < \frac{900 - 12\mu}{34.64}\right] = 0.9 \rightarrow \frac{900 - 12\mu}{34.64} = 1.28 \rightarrow \mu = 71.30 \;\; kilo$$

So individual weight on average must at most 71.30 to hold the stated condition.

[124] In a production process the probability of a faulty item is 0.25. We have a 100 item batch.
(a) What is the probability of having 30 faulty items or less?
(b) How many faulty items are expected to be?
(c) What is the probability of having exactly the number of faulty items expected? Interpret the result.
(d) How many faulty items can we ensure with a probability of 90%?

(a) X number of faulty items into the batch follows a binomial distribution with large n, so it can be approximate by the normal distribution:

$$X \sim B(n = 100, p = 0.25) \rightarrow N(\mu = np = 25, \sigma = \sqrt{npq} = \sqrt{np(1-p)} = 4.33)$$

So, after fixing the lag between the no. of faulty items as a discrete value and the normal distribution as a continuous distribution:

$$P[X \leq 30] = P[X < 30.5] = P\left[Z < \frac{30-25}{4.33}\right] = P[Z < 1.15] = 0.87$$

(b) There are expected to be $\mu = np = 25$ faulty items.

(c)

The expected value is $\mu = 25$.

After fixing the discrete/continuous lag (we say that there are 25 faulties whien on the continuous interval no. or *amount* of faulties is between 24.5 and 25.5:

$$P[X = 25] = P[24.5 < X < 25.5] = P\left[\frac{24.5-25}{4.33} < Z < \frac{25.5-25}{4.33}\right] = P[-0.11 < Z < 0.11] = 0.087$$

It's also the value with the largest probability, but it we can see, its probability is not so big. That's because it's an isolated value. Big probabilities will be for a wider interval around μ.

(d)

Ensuring a number of faulty items is giving a maximum number for them, otherwise it's no sense. Hence,

$$P[X < x] = 0.9 \rightarrow P\left[Z < \frac{x-25}{4.33}\right] = 0.9 \rightarrow \frac{x-25}{4.33} = 1.28 \rightarrow x = 30.54$$

In this case we must take the highest integer, that is 31, in order to have the probability above 0.9.

[125] 1000 people are attending a recruitment process. The probability of passing the first exam is 0.6. How many chairs should we get if we want the probability of all the candidates in the second exam having a chair available to be at least 0.99?

The number of peole passing the first exam is this one:

$$X \sim B(n = 1000, p = 0.6) \to N(\mu = 1000 \times 0.6 = 600, \sigma = \sqrt{1000 \times 0.6 \times 0.4} = 15.49)$$

To solve this kind of problem (looking for a value that gives an exact probability) is very useful to try wuth a given value. As on average $1000 \times 0.6 = 600$ people pass the first exam, on average we will need 600 chairs for those people on the second exam. As we want a big probability, 0.99. of having enough chairs, we should try with a number chairs larger than 600. Le'ts try with, for example, 700 chairs:

$$P[enough] = P[X < 700] = P\left[Z < \frac{700 - 600}{15.49}\right] = P[Z < 6.45] \approx 1$$

So, we have enough (more than enough) with 700 chairs. In order to get the exact value for 0.99 probability:

$$P[enough] = P[X < x] = P\left[Z < \frac{x - 600}{15.49}\right] = 0.99$$

We lok for the z socre into the tables:

$$\frac{x - 600}{15.49} = 2.32 \to x = 635.93$$

The number of chairs must be integer. With 635 chairs we will be short, so we will need 636 chairs.

---

[126] Daily production follows a N(100kg, 10kg) distribution. What is the probability of having in one year (365 days) at least 317 days with a production of at most 110kg?

Let's calculate the probability of producing at most 110kg in one day:

$$P[X < 110] = P\left[Z < \frac{110 - 100}{10}\right] = P[Z < 1] = 0.8413$$

So D number of days within a year with a production level smaller than 110 kg distributes in this manner:

$$D \sim B(n = 365, p = 0.8413) \to N(\mu = np = 365 \times 0.8413 = 307.07, \sigma = \sqrt{npq} = 6.98)$$

And the probability required:

$$P[D > 317] = P\left[Z > \frac{317 - 307.07}{6.98}\right] = P[Z > 1.42] = 0.077$$

[127] The expected daily number of failures in a machine is 0.16, following the Poisson distribution. We need one piece to fix every failure.

i. What is the probability of having more than 50 failures in a year?

ii. How many pieces do we need in order to fix all the failures in a year with a 0.99 probability?

iii. With 100 pieces, for how many days can we ensure that we have enough pieces to fix all the failures with a 0.95 probability?

(a)

Let's calculate lambda parameter for one year: $\lambda_{1\ year} = 365 \times 0.16 = 57.6$

As lambda is bigger than 30, we can use the normal approximation to calculate probabilities:

$$P(\lambda = 57.6) \to N(\mu = \lambda = 57.6, \sigma = \sqrt{\lambda} = 7.59)$$

Let's calculate finally the requested probability:

$$P[X > 50] = \text{(continuity correction)} = P\left[Z > \frac{50.5 - 57.6}{7.59}\right] = P[Z > -0.93] = P[Z < 0.93] = 0.8238$$

(b)

When we have x pieces, they will be enough when failures are x or lesser (we won't apply the continuity correction in this case, the error is very small):

$$P[enough] = P[X < x] = P\left[Z < \frac{x - 57.6}{7.59}\right] = 0.99 \to \frac{x - 57.6}{7.59} = 2.32 \to x = 75.20$$

With 75 pieces we are short, so we take 76 pieces to fulfill the requested goal.

(c)

When we have 100 pieces, they will be enough when failures are 100 or less. In this case, we don't know the numer fo days n, so the number of failures distribute in this way:

$$P(\lambda = 0.16n) \to N(\mu = \lambda = 0.16n, \sigma = \sqrt{0.16n} = 0.4\sqrt{n})$$

$$P[enough] = P[X < 100] = P\left[Z < \frac{100 - 0.16n}{0.4\sqrt{n}}\right] = 0.99 \to \frac{100 - 0.16n}{0.4\sqrt{n}} = 1.64 \to n = 530.38$$

So we will have enough for 530 days (531 days is too much).

[128.] Following a Poisson process, 1.4 batteries are exhausted on average each day in a machine.

(a) Throughout 40 days, what is the probability of exhausting more than 50 batteries?

(b) How many batteries should we get in order to have enough energy for 80 days with a 0.9 probability?

(a)

We must adjust lamnbda to a period of 40 days:

$$\lambda_{1\ day} = 1.4 \rightarrow \lambda_{40\ days} = 40 \times 1.4 = 56$$

As lambda is big enough, we may use the normal distribution to calculate probabilitites (X:exhausted batteries):

$$P(\lambda = 56) \rightarrow N(\mu = 56, \sigma = \sqrt{56} = 7.4)$$

$$P[X > 50] = P\left[Z > \frac{50 - 56}{7.4}\right] = P[Z > 0.81] = 0.2089701$$

(b)

We must adjust lambda to a period of 80 days:

$$\lambda_{1\ day} = 1.4 \rightarrow \lambda_{80\ days} = 80 \times 1.4 = 112$$

On average we need 112 batteries, but as we need bigger certainty about having enough, we will need a big ger amount of them. Let's take 120 (we don't perform the ± 0.5 correction):

$$P(\lambda = 112) \rightarrow N(\mu = 112, \sigma = \sqrt{112} = 10.583)$$

$$P[enough] = P[X \leq 120] = P\left[Z < \frac{120 - 112}{10.583}\right] = P[Z < 0.75] = 0.7733726$$

This probability is smaller that our goals's, so we need more batteries than 112. Going backwards:

$$0.98 = P\left[Z < \frac{x - 112}{10.583}\right] \rightarrow \frac{x - 112}{10.583} = 2.05 \rightarrow x = 133.69$$

133 is short for us, so by caution we will take 134 batteries.

[129] Substance consumed in a day follows an uniform distribution, in a 5 to 10 liters interval.

i. What is the probability of consuming less than 420 liters in 40 days?

ii. How many liters must we get for 40 days in order to being the probability of having enough substance 0.99?

iii. Idem, if the consumption follows a U(0,15) distribution.

iv. Provided we have 500 liters, for how many days do we have enough substance with a 0.99 probability?

(a) Consumption along 40 days is calculated in this way:

$$X = X_1 + X_2 + \cdots + X_{40}$$

X_i daily consumptions distribute $U(5,10)$ and therefore mean and variance for each are these:

$$\mu = \frac{5+10}{2} = 7.5,\ \sigma^2 = \frac{(10-5)^2}{12} = 2.08$$

Assuming independence, we can apply CLT as we have an enough no. of adding up distributions. Hence:

$$X \sim N(\mu = 7.5 + 7.5 + \cdots + 7.5 = 40 \times 7.5 = 300; \sigma = \sqrt{40 \times 2.08} = 9.12)$$

Let's calculate the requested probability:

$$P[\mathbf{X} > 420] = P\left[Z > \frac{420-300}{9.12}\right] \approx 0$$

(b) We will have enough substance with x liters (unknown) when consumption along 40 days is smaller than x. Hence:

$$P[X < x] = P\left[Z > \frac{x-300}{9.12}\right] = 0.99 \rightarrow \frac{x-300}{9.12} = 2.32 \rightarrow x = 318.24l$$

(c) For a daily consumption folowing a $U(0,15)$ distribution, we have:

$$\mu = \frac{0+15}{2} = 7.5,\ \sigma^2 = \frac{(15-0)^2}{12} = 18.75$$

Remark than the mean is the same and the deviation is bigger. So, now: $X \sim N(\mu = 300; \sigma = \sqrt{40 \times 18.75} = 27.38)$

And developing just as in the previous section:

$$P[X < x] = P\left[Z > \frac{x-300}{27.38}\right] = 0.99 \rightarrow \frac{x-300}{27.38} = 2.32 \rightarrow x = 354.7l$$

We need more liters. Why? Mean is the same but deviation (spread of possible values or uncertainty) is bigger. So we need more liters to have the same probability of having enough.

(d)

Consumption along n days (unknown) distributes:

$$X \sim N(\mu = 40n; \sigma = \sqrt{40n} = 6.32\sqrt{n})$$

We have enough with 500l when total consumption is smaller than 500.

$$P[enough] = P[X < 500] = P\left[Z > \frac{500 - 40n}{6.32\sqrt{n}}\right] = 0.99 \rightarrow \frac{500 - 40n}{6.32\sqrt{n}} = 2.32$$

Solving the resulting quadratic equation ($x = \sqrt{n}, x^2 = n$):

$$x = 3.36 \rightarrow n = 11.28 \rightarrow 11 \; days$$

We take 11 days, because along 12 days we will have scarcity with 500l.

12 is smaller than 30 (no. of dist. to add up needed to apply CLT), but in the special case of uniform with $n = 12$ we can apply CLT.

[130] The share price possible daily increments are +1, 0 and -1 with respectively 0.2, 0.5 and 0.3 probabilities. (a) After 100 days, what is the probability of not losing? (b) How much money should we have after 100 days in order to be able to pay the losses with a 0.99 probability?

(a) We know about daily increments or benefits but they are asking us about benefits throughout 100 days. Let's link them:

$$\mathbf{X}_{100\ days} = X_1 + X_2 + \ldots + X_{100}$$

As we know the distribution for each summand, they are independent and have a big enough number of them ($n > 30$), we can apply CLT for the sum. To apply CLT we need the mean and variance of each summand:

x	$p(x)$	$xp(x)$	$x^2p(x)$
-1	0.3	-0.3	0.3
0	0.5	0	0
1	0.2	0.2	0.2
	1	-0.1	0.5

So, for each day:

$$\mu = -0.1\ ;\ \sigma^2 = 0.5 - (-0.1)^2 = 0.49$$

Hence, applying CLT for 100 days:

$$\mathbf{X}_{100\ days} \sim N(\mu = -0.1 \times 100 = -10, \sigma = \sqrt{0.49 \times 100} = 7)$$

Now we can calculate the probability:

$$P[not\ losing] = P[\mathbf{X}_{100\ days} > 0] = P\left[Z > \frac{0 - (-10)}{7}\right] = P[Z > 1.42] = 0.077$$

(b)

On average we will have losses of 10 euros. As we want to be almost sure that we will have enough money to pay all losses, logically the solution will be bigger than 10 euros. Let's take for example 11 euros. 11 euros will enough when benefits are larger than -11:

$$P[enough] = P[X > -11] = P\left[Z > \frac{-11 - (-10)}{7}\right] = P[Z > -0.14] = 0.55$$

With 11 euros we don't reach the goal, so we need more money. Let's call that amount of money m. As we want the probability to be 0.99, going back from the last expression:

$$P[enough] = P[X > -m] = P\left[Z > \frac{-m - (-10)}{7}\right] = P\left[Z > \frac{-m + 10}{7}\right] = 0.99 \rightarrow \frac{-m + 10}{7} = -2.32 \rightarrow m = 26.24$$

So, the amount of money to finance possible losses with a 0.99 probability will 26.24 euros.

[132] In a factory the daily average production from Monday to Thursday is 146Tn, with a standard deviation of 10Tn, following an unknown distribution. On Fridays and Saturdays, the factory works only along the morning, and so the average production is 64Tn with a standard deviation of 6Tn.

(a) If along the next 8 weeks we are working only from Monday to Thursday, how much production can we guarantee for those days with a 0.96 probability?

(b) Calculate the probability of less than 7000 tons along 10 weeks.

(c) Give the deadline or number of weeks needed to produce 8000 tons with a 0.96 probability?

(a)

Alog the next 8 weeks we have 32 days, with an 146 average and a 10 standard deviation. Let's give how the production distributes along those 32 days.

Production along 32 days is the sum of productionf along 32 days:

$$\mathbf{X} = X_1 + X_2 + \cdots + X_{32}$$

Production for each day distributes in this way (unknown exact distribution, but known parameters):

$$X_i \sim Dist?(\mu = 146, \sigma = 10)$$

Hence, assuming independence amog production of different days and as the number of days (or distributions) we add is bigger than 30, we can use CLT (remark that we don't need the exat distribution of the production for each day, it's enough with them mean and the standard deviation)

$$\mathbf{X} \sim N(\mu = 146 \times 30 = 4672, \sigma = \sqrt{10^2 \times 32} = 56.56)$$

Guaranteeing a production level is giving a minimum for it:

$$P[\mathbf{X} > x] = 0.96;\ x?$$

Standardizing:

$$P[\mathbf{X} > x] = P\left[Z > \frac{x - 4672}{56.56}\right] = 0.96$$

z score must be negative, in order to leave above it a 0.96 probability:

$$\frac{x - 4672}{56.56} = -1.75$$

Hence: $x = 4573.02 Tons$

(b)

Along the next 10 weeks we have 40 days from Monday to Thursday (X days), and 20 days from Friday to Saturday (Y days).

Total production is the sum of productions of those 60 days:

$$\mathbf{X} = X_1 + X_2 + \ldots + X_{40} + Y_1 + Y_2 + \ldots + Y_{20}$$

Applying CLT, assumed independence:

$$\mathbf{X} \sim N(\mu = 146 \times 40 + 64 \times 20 = 7120; \sigma = \sqrt{10^2 \times 40 + 6^2 \times 20} = 68.7)$$

Hence: $P[\mathbf{X} < 7000] = P\left[Z < \frac{7000 - 7120}{68.7}\right] = 0.04$

(c)

Along the deadline, that is to say in a weeks, we have X type $4a$ days, and Y type $2a$ days. So, by CLT, the total production till the deadline distributes in this way:

$$\mathbf{X} \sim N(\mu = 146 \times 4a + 64 \times 2a = 712a, \sigma = \sqrt{10^2 \times 4a + 6^2 \times 2a} = 21.72\sqrt{a})$$

The deadline will be fulfilled if along a weeks the totoal production is 8000 or larger. As we want that probability to be at least 0.96:

$$P[\mathbf{X} \geq 8000] = 0.96$$

Standardizing:

$$P[\mathbf{X} > 8000] = P\left[Z > \frac{8000 - 712a}{21.72\sqrt{a}}\right] = 0.96$$

z score leaving above it a 0.94 probability is -0.75, looking into the tables and applying symmetry:

$$\frac{8000 - 712a}{21.72\sqrt{a}} = -1.75$$

We order the equation:

$$712a - 38\sqrt{a} - 8000 = 0$$

With $n = x^2$ transform, we get a quadratic equation:

$$712x^2 - 38x - 8000 = 0$$

We get the positive root (the negative one gives a negative sigma, and that's not possible):

$$x = 3.37 \rightarrow a = x^2 = 11.35$$

a must be an intger number. We take 12 weeks, as with 11 weeks we are short. So, we have to set a 12 weeks deadline to keep the goal.

[134] In a given flight, we assume that the weight of each baggage is on average 20kg, with a 3kg deviation. The exact distribution is unknown.

(a) If the flight gets 100 travelers, what is the probability of having more than 2100kg baggages?

(b) Which amount of baggages can you ensure with 96% probability?

(c) How many travelers can we take with the condition that not more of 2000kg will be taken with a 0.99 probability?

(d) Calculate the probability of not having more than 100kg with 5 baggages. Why can't you apply CLT in ths case? What should you apply? And what if weight of baggages were normal distributed?

(e) What is the probability of the sample mean of 50 baggages being bigger than 21?

(f) If the sample mean of 50 baggages was really bigger than 21kg, which should your decision about the theoretical average weight? Significance level: 0.05

(d)

In this case we cannot apply CLT, because the number of adding up distributions is too small. Nevertheless, the mean weight for 5 baggages will be $\mu = 5 \times 20 = 100$, and the standard deviation $\sigma = \sqrt{5 \times 3^2}$. Known the mean and the standard deviation, we can give a bound for any probability by Chebyshev's inequality (in a previous lesson). If weights were normal distributed we could apply the property of reproductivity for normal distributions and apply it, just like in previous problems (see section about normal distribution).

(e)

The sample mean of 50 baggages (the sum of 50 baggages taken randomly divided by 50) distributes this way by CLT:

$$\overline{x} \sim N\left(\mu = 20, \sigma = \frac{3}{\sqrt{50}} = 0.42\right)$$

So the probability requested is:

$$P[\overline{x} > 21] = P\left[Z > \frac{21 - 20}{0.42}\right] = P[Z > 2.38] = 0.008$$

(f)

With $\overline{x} = 21$, it looks like the mean weight has increased, so we take as the null hypothesis the opposite:

$$H_0 : \mu \leq 20$$

The test direction is upwards: we reject the null hypothesis about the theoretical mean weight (better, population mean) when the sample mean is big. So we must calculate $P[\overline{x} > 21]$.

As the previous probability is smaller than α, we reject the null hypothesis and state that the evidence is strong enough to state that the population mean (the mean for all baggages) has increased.

[135] Daily production in a factory follows a U(10,20) distribution, in kilograms. In 50 days, the average production has been 140 kg.
(a) Should we conclude that production has decreased on average? Significance level: 2%.
(b) Which is the production level in order to claim that average production has decreased? Significance level: 2%.
(c) Calculate the critical values in order to reject the null hypothesis with sample sizes of 100 and 500 days. Interpret the results.
(d) Which is the production level in order to claim that average production has just changed? Significance level: 2%.

(a)

We have to perform a statistical test. So we have to calculate a probability about evidence, in this case a sample mean. For that, we have to set the distribution about the sample mean for 50 days (X_i: daily production):

$$\overline{x}_{n=50} = \frac{\sum_{i=1}^{50} X_i}{50} = \frac{X_1 + X_2 + \cdots + X_{50}}{50}$$

We can apply CLT to $sum_{i=1}^{50} X_i$, as number of summands is big enough ($n > 30$), and those are independent between them. To Apply CLT we have to calculate expectation and variance for each summand (daily production):

$$X \sim U(10,20) \begin{cases} \mu = \dfrac{10+20}{2} = 15 \\ \sigma^2 = \dfrac{(20-10)^2}{12} = 8.33 \end{cases}$$

Applying CLT now:

$$\sum_{i=1}^{50} X_i \sim N(\mu = 15 \times 50, \sigma = \sqrt{50 \times 8.33})$$

Applying linear transformation for normal distributions:

$$\overline{x}_{n=50} \sim N\left(\mu = \frac{15 \times 50}{50} = 15; \sigma = \frac{\sqrt{50 \times 8.33}}{50} = \frac{\sqrt{8.33}}{\sqrt{50}} = 0.408\right)$$

Sample mean is 14, but on average is 15. So, as it seems it has decreased, by caution we take as null hypothesis the opposite (things have not changed):

$$H_0 : \mu = 15$$

As we are amazed because sample mean is small, we calculate the lower probability:

$$P[\overline{x}_{n=50} < 14] = P\left[Z < \frac{14-15}{0.408}\right] = P[Z < -2.45] = 0.007 < \alpha$$

Hence, we reject the null hypothesis and so there is enough evidence to state that mean has decreased.

The probability for evidence (or something larger) is called *p-value* in statistics.

(b)

To claim that the mean has decreased, p-value must be smaller than 0.02:

$$P[\overline{x}_{n=50} < \overline{x}^*_{n=50}] = P\left[Z < \frac{\overline{x}^*_{n=50} - 15}{0.408}\right] = 0.02$$

The standard score that leaves below it a 0.02 probability is -2.05.

$$\frac{\overline{x}^*_{n=50} - 15}{0.408} = -2.05 \rightarrow \overline{x}^*_{n=50} = 14.16$$

So 50 days sample mean must be lower than 14.16 to state that population mean has decreased from 15. We call the 14.16 value the *critical value* and it's the boundary between acceptance and rejection region. We will also call reject region *critical region*. As 14 is on the rejection region, we reject the null hypothesis.

Both methods to solve the statistical test, p-value and critical region, lead always to the same result.

(c) Changing sample size (number of days, in this case) keeps the mean for the sample mean, but changes the standard deviation:

$$\overline{x}_{n=100} \sim N\left(\mu = 15; \sigma = \frac{\sqrt{8.33}}{\sqrt{100}} = 0.288\right)$$

$$\overline{x}_{n=500} \sim N\left(\mu = 15; \sigma = \frac{\sqrt{8.33}}{\sqrt{500}} = 0.129\right)$$

Let's calculate now the critical values:

$$\frac{\overline{x}^*_{n=100} - 15}{0.288} = -2.05 \rightarrow \overline{x}^*_{n=100} = 14.40$$

$$\frac{\overline{x}^*_{n=500} - 15}{0.129} = -2.05 \rightarrow \overline{x}^*_{n=500} = 14.73$$

As it can be seen, the bigger the sample size is, the easier will it be to reject the null hypothesis: a given deviation from the 15 value will be more and more significant, as we have more information with a larger sample size.

(d) To claim the population mean (15) has just changed, sample mean must be very big or very small. So, rejection region will be both in the lower and the upper side. So, we will say in that case that test is two-tailed or two-sided, in contrast with one-tailed or one-sided tests in the previous sections of this problem.

To perform a two-sided test, we have to divide the significance level between both tails, 1% for each one.

Now, we calculate the critical values:

- **Lower side:**

$$P[\overline{x}_{n=50} < \overline{x}^*_{n=50}] = P\left[Z < \frac{\overline{x}^*_{n=50} - 15}{0.408}\right] = 0.01$$

The standard score leaving a 0.01 probability *below* is -2.32.

$$\frac{\overline{x}^*_{n=50} - 15}{0.408} = -2.32 \rightarrow \overline{x}^*_{n=50} = 14.05$$

- **Upper side:**

$$P[\overline{x}_{n=50} > \overline{x}^*_{n=50}] = P\left[Z > \frac{\overline{x}^*_{n=50} - 15}{0.408}\right] = 0.01$$

The standard score leaving a 0.01 probability *above* is 2.32.

$$\frac{\overline{x}^*_{n=50} - 15}{0.408} = 2.32 \rightarrow \overline{x}^*_{n=50} = 15.95$$

Hence, we will reject the mean daily production of 15 when 50 days sample mean is lower tha 14.05 or larger than 15.95.

[140] A restaurant serves menus from Monday to Friday, at noon and evenings. The total number of menus served is given below:

Menus	Monday	Tuesday	Wednesday	Thursday	Friday	Total
Midday	**38**	**45**	**38**	**58**	**40**	**219**
Evening	**18**	**31**	**26**	**30**	**36**	**141**
Total	**56**	**76**	**64**	**88**	**76**	**360**

i. Conclude whether number of menus served from Monday to Friday follows an uniform distribution, by means of the chi-square test. Significance level: 5%.

ii. Test whether menus served at noon are twice menus served at evening. Significance level: 5%.

(a)

Week day	Observed: O_i	Probability: p_i	Expected: E_i	$\frac{(O_i-E_i)^2}{E_i}$
Monday	56	0.2	72	3.55
Tuesday	76	0.2	72	0.22
Wednesday	64	0.2	72	0.88
Thursday	88	0.2	72	3.55
Friday	76	0.2	72	0.22
Total	360	1	360	8.42

$$\mathbf{X}^2 = 8.42 \ ; \ \chi^2_{5-1,0.95} = 9.49$$

As chi-square statistic is lower than the critical value, we conclude that the difference between observed and expected frequencies, given by chi square statistic, is not significant. Hence, we accept H_0 and conclude the model is correct, so we may state for those data that distribution of menus along week days is uniform.

(b)

Day period	Observed: O_i	Probability: p_i	Expected: E_i	$\frac{(O_i-E_i)^2}{E_i}$
Noon	219	2/3	240	1.83
Evening	141	1/3	120	3.67
Total	360	1	360	5.51

$$\mathbf{X}^2 = 5.51 \ ; \ \chi^2_{2-1,0.95} = 3.84$$

As chi-square statistic is larger than the critical value, we conclude it's significant. Hence, we reject H_0 and conclude the model is wrong, so we cannot state for those data that distribution of menus along day is given by the 2/1 proportion.

---

[141] Times in days till a failure in a machine are given below:

$26.2, 22.3, 33.5, 19.0, 24.7, 25.6, 26.2, 28.9, 27.6, 26.5, 27.1$

$32.4, 36.2, 34.1, 28.7, 26.5, 25.4, 23.4, 21.6, 22.0, 20.6, 30.2$

Setting < 19, $19-21$, $21-23$, ... intervals, test whether data follow a normal distribution, by means of the chi-square test, provided that we must previously estimate the mean and the standard deviation. Significance level: 10%.

Same as 115. problem, but here we have to set the intervals and make the count for the number of data in each one. We can calculate the estimates for μ and σ fron raw data or interval data (the latter like in the 115. problem). In order to avoid the interval approximation error, the best is to take the raw data to make the estimates:

$$\hat{\mu} = \overline{x} = \frac{26.2 + 22.3 + \cdots}{22} = 26.76$$

$$\hat{\sigma} = \hat{s} = \sqrt{\frac{(26.2 - 26.76)^2 + (22.3 - 26.76)^2 + \cdots}{22 - 1}} = 4.51$$

The ongoing method is the same as in 115. problem. As we make 2 estimates, degrees of freedom should be *number of intervals-1-2* (the number of intervals depends on the intervals we take on the upper side of data).

[142] Times in days till a failure in a machine are given below:

$26.2, 22.3, 33.5, 19.0, 24.7, 25.6, 26.2, 28.9, 27.6, 26.5, 27.1$

$32.4, 36.2, 34.1, 28.7, 26.5, 25.4, 23.4, 21.6, 22.0, 20.6, 30.2$

Setting $0-10$, $10-20$, $20-30$, $30-40$ and >40 intervals, test whether data follow an exponential distribution, by means of the chi-square test, provided that we must previously estimate the mean of the distribution. Significance level: 10%.

Estimate about the mean. (Remember: mean for the exponential distribution is $\frac{1}{\lambda}$.)

$$\frac{1}{\hat{\lambda}} = \overline{x} = 26.76 \rightarrow \hat{\lambda} = 0.037$$

Theoretical probabilities:

$$P(X<10) = 1-e^{-0.037\times 10} = 0.31$$
$$P(10<X<20) = P(X<20)-P(X<10) = [1-e^{-0.037\times 20}]-[1-e^{-0.037\times 10}] = 0.52-0.31 = 0.21$$
$$P(20<X<30) = P(X<30)-P(X<20) = [1-e^{-0.037\times 30}]-[1-e^{-0.037\times 20}] = 0.67-0.52 = 0.15$$
$$P(30<X<40) = P(X<40)-P(X<30) = [1-e^{-0.037\times 40}]-[1-e^{-0.037\times 30}] = 0.77-0.67 = 0.10$$
$$P(X>40) = 1-[1-e^{-0.037\times 40}] = 0.23$$

Chi-square calculation:

Intervals	Observed: O_i	Prob.: p_i	expected: E_i	$\frac{(O_i-E_i)^2}{E_i}$
0-10	0	0.31	6.82	6.82
10-20	1	0.21	4.62	2.83
20-30	16	0.15	3.3	48.87
30-40	5	0.10	2.2	3.56
>40	0	0.23	5.06	5.06
Total	22	1	22	67.14

$$\mathbf{X}^2 = 67.14\ ;\ \chi^2_{5-1-1,0.9} = 6.25$$

Chi-square statistic is significant, and so it is difference between observed and expected frequencies. So we reject the model and state that exponential distribution is not fit for those data.

[147] We have compiled some invoice amounts paid in a shop by sex:

$Men : 3, 3, 5, 6, 8, 10, 10, 11, 11, 12, 12, 12, 16, 19, 20$

$Women : 2, 7, 9, 11, 13, 13, 15, 17, 17, 18, 20, 21, 23, 24, 25, 25, 27, 32, 36, 39$

Test with those data whether men and women have the same buying behavior. Hint: 1+2+...+35=630. Significance level: 5%. Solve it looking into the corresponding tables as well as by the normal approximation.

---

x_{ord}	Sex	Rank
2	W	1
3	M	2
3	M	3
5	M	4
6	M	5
7	W	6
8	M	7
9	W	8
10	M	9.5
10	M	9.5
11	M	12
11	M	12
11	W	12
12	M	15
12	M	15
12	M	15
13	W	17.5
13	W	17.5
15	W	19
16	M	20
17	W	21.5
17	W	21.5
18	W	23
19	M	24
20	M	25.5
20	W	25.5
21	W	27
23	W	28
24	W	29
25	W	30.5
25	W	30.5
27	W	32
32	W	33
36	W	34
39	W	35

$$W_{men} = 178.5$$

No need to calculate the sum of ranks for women: the sum of all ranks (men and women together) is the sum of the first 35 natural numbers (use the formula for an arithmetic progression, or take directly the hint given in the problem statement):

$$S = 35 \times \frac{1+35}{2} = 630 \rightarrow W_{women} = 630 - 178.5 = 451.5$$

The minimum is:

$$W_{min} = 178.5$$

Looking into the tables for a two-sided test ($n_{men} = 15$, this gives the minimum), we see that the critical value is 210. As W_{min} statistics is lower than the critical value we reject the null hypothesis and state that men and women are different and cannot be put together into one data set.

Using the normal approximation:

$$W_{min} \sim N\left(\mu = \frac{15 \times (15+24+1)}{2},\ \sigma = \sqrt{\frac{15 \times 24 \times (15+24+1)}{12}}\right)$$

$$W_{min} \sim N(\mu = 300; \sigma = 34.64)$$

Taking W_{min}, the critical region (the amazing thing) is always on the lower side. So, we calculate in this manner the p-value:

$$P[W_{min} < 178.5] = P\left[Z < \frac{178.5 - 300}{34.64}\right] = P[Z < -3.53] = 0.0002 < 0.025$$

It's a two-sided test, so we have to compare the p-value to $\alpha/2 = 0.025$.
The conclusion is to reject the null hypothesis, so we state that men and women are different. Hence, we cannot put all of them into one data set, and we must study them apart.
We may also solve the test by calculating the W^* critical value:

$$P[W_{min} < W^*] = P\left[Z < \frac{W^* - 300}{34.64}\right] = 0.025 \rightarrow \frac{W^* - 300}{34.64} = -1.96 \rightarrow W^* = 232.1$$

As the value for W_{min} statistic is lower than the critical value, it's on the rejection region. So we reject the homogeneity, and state that men and women are different.

[148] Children with math comprehension problems were given a special training program last year. They carried a test before and after the program. The results are given below:

$Before: 22, 32, 43, 28, 27, 36$

$After: 25, 42, 50, 35, 35, 42$

Test whether the program has been successful by means of the Wilcoxon rank sum test. Significance level: 5%. Hint: 1+2+...+12=78. Remark: surveyed children before and after the program are different , so we have independent samples. If the children were the same, we would have dependent samples (paired samples) and hence we would have to perform another kind of test (sign test, for example).

---

x_{ord}	Group	Rank
22	before	1
25	after	2
27	before	3
28	before	4
32	before	5
35	after	6
35	after	7
36	before	8
42	after	9
42	after	10
42	before	11
50	after	12

As ties are from the same group, it's not worthwhile to solve them taking the middle point:

$$W_{before} = 32\ ;\ W_{after} = 78 - 32 = 46$$

We state that the program has been successful when W_{after} is big enough, that is, when W_{before} is small enough. On the other hand, W_{before} is the statistic giving the minimum sum of ranks. So, looking in to the tables for $n = m = 6$ on one-sided tests, the critical value is 28. As $W_{before} = 32$ is bigger, we accept the hypothesis of homogeneity, so we cannot state at that significance level that the program has been successful.

[149] Data about number of movie tickets sold on Saturday and Sunday are given below:

Saturday : 126 – 91 – 68 – 122 – 113 – 137 – 111 – 86 – 100 – 82 – 96 – 121 – 97 – 95 – 89

Sunday : 81 – 98 – 129 – 101 – 121 – 124 – 133 – 108 – 84 – 89 – 86 – 131

Test if more tickets are sold on Sunday, by means of Wilcoxon rank-sum test and using the normal approximation as well as the tables. Significance level: 1%. Hint: 1+2+...+27=378.

It's not a two-sided test, as the decision is towards a given direction for the decision (more on Sundays?).

68 Sat 1	81 Sun 2	82 Sat 3	84 Sun 4	86 Sat 5	86 Sun 6	89 Sun 7.5	89 Sat 7.5	91 Sat 9
95 Sat 10	96 Sat 11	97 Sat 12	98 Sun 13	100 Sat 14	101 Sun 15	108 Sun 16	111 Sat 17	113 Sat 18
121 Sat 19.5	121 Sun 19.5	122 Sat 21	124 Sun 22	126 Sat 23	129 Sun 24	131 Sun 25	133 Sun 26	137 Sat 27

$n_{sat} = 15;\ W_{sat} = 198$

$n_{sun} = 12;\ W_{sun} = 180$

Taking W_{sun} as a basis:

by means of p-value

$$W_{sun} \sim N\left(\mu = \frac{12 \times (12+15+1)}{2}, \sigma = \sqrt{\frac{12 \times 15 \times (12+15+1)}{12}}\right) : N(\mu = 168, \sigma = 20.49)$$

We will reject homogeneity and accept we sell more on Sundays when W_{sun} is big enough:

$$P[W_{sun} > 180] = P\left[Z > \frac{180-168}{20.49}\right] = P[Z > 0.58] = 0.28 > \alpha$$

We compare to α as it's a one-sided test.

So, we accept homogeneity (we don't sell more on Sundays).

by means of critical value or critical region (rejection region)

$$P[W_{sun} > W^*] = P\left[Z > \frac{W^*-168}{20.49}\right] = 0.01 \rightarrow \frac{W^*-168}{20.49} = 2.32 \rightarrow W^* = 215,53$$

Actually, we reject homogeneity and state that sales are bigger on Sundays when W on Sundays is *big*. So rejection region is on the upper side from the critical value. $W_{sun} = 180$ is below it, so we accept homogeneity (not possible to state we sell more on Sundays).

Looking into the tables

We reject homogeneity and state that we sell more on Sundays when W_{sun} is big and W_{sat} *is small*. As into the tables we look for the W_{min} value or towards the lower side, we must compare the critical value to W_{sat}, and not to W_{sun}. The critical value for one-sided tests is 120 for $n = 12, m = 15$ (n being the sample size of the W we take as the reference value). As $W_{sat} = 198$ is bigger than $W^* = 120$, we must accept homogeneity and state that there's no evidence to sell more on Sundays.

[154] We have drawn 50 data about time following an alleged exponential distribution. Normally mean time is 100 minutes. Set the test in order to conclude whether mean time has decreased. Significance level: 5%.

We have not drawn any data. So we have no choice: we can only design the test, giving the critical value for the rejection region.

As we have to decide if the mean has decreased, we take the opposite in the null hypothesis, cautiously (2nd criterion):

$$H_0 : \mu = \frac{1}{\lambda} \geq 100$$

Deviation is not given but can be drawn from the mean (this is a hard point! remember the model is exponential!):

$$\lambda = \frac{1}{100} \rightarrow \sigma^2 = \frac{1}{\lambda^2} = 10000 \rightarrow \sigma = \sqrt{10.000} = 100$$

Sample mean (we take it as the evidence for μ) distributes in this manner:

$$\overline{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

So we have:

$$\overline{x} \sim N\left(100, \frac{\sqrt{10000}}{\sqrt{50}} = 14.14\right)$$

We will reject that $\mu > 100$ when $\overline{x}$ is small. So the rejection region is on the left (or lower) side:

$$P[\overline{x} < \overline{x}_0] = P\left[Z < \frac{\overline{x}_0 - 100}{14.14}\right] = 0.05 \rightarrow \frac{\overline{x}_0 - 100}{14.14} = -1.64 \rightarrow \overline{x}_0 = 76.81$$

We will reject H_0 when the sample mean is lesser than 76.81. We have no data but we have designed the test.

[155] We have compiled water comsumption data from 80 families: $\overline{x} = 7$; $\sum x_i^2 = 4800$. The model for data is supposed to be a gamma distribution, with unknown mean and variance.

- **i. Decide by means of the p-value if population mean may be bigger than 7.5. Significance level: 1%.**
- **ii. Now decide if population mean may be smaller than 7.5. Significance level: 1%.**

(a) We take as the null hupothesis the opposite of the statement we want to prove. So:

$$H_0 : \mu \leq 7.5$$

We reject the null hypothesis when the sample mean ($\overline{x}$) is bigger enough than 7.5. Here we have, $\overline{x} = 7$, so it's quite clear that the null hypothesis will be accepted (there is no reason to reject it!).

The standard deviation of the population (σ) is unknown, so we must estimate it from data:

$$s^2 = \frac{4300}{80} - 7^2 = 4.75 \to \hat{s}^2 = \frac{80}{79} \times 4.75 = 4.81 \to \hat{s} = 2.19$$

In this context (non-normal model, big sample) the sample mean distributes in this way:

$$\overline{x} \sim N\left(\mu, \frac{\hat{s}}{\sqrt{n}}\right) : N\Big(7.5, \frac{2.19}{\sqrt{80}}\Big) : N(7.5, 0.24)$$

The test is 1-sided. We reject $H_0 : \mu \leq 7.5$ when the sample is very bih. Let's calculate the p-value in that direction:

$$P[\overline{x} > 7] = P\Big[Z > \frac{7 - 7.5}{0.24}\Big] = P[Z > -2.08] = 0.9812 > \alpha$$

The p-value is (much) bigger than the significance level, so we accept the null hypothesis (loosely). There is no reason to state that the mean is bigger than 7.5, as we advanced.

[156] A component's duration follows a normal distribution, with a standard deviation of 4 days. Quality specifications require mean duration to be at least 10 days. Sample mean from 9 data has been 8 days. Test whether the specification is met or not. Significance level: 1%.

It's not very clear what we want to decide: whether the specification is met or not. So, we cannot clearly set a null hypothesis following the [2] criterium.

On the other side, taking the [3] criterium, sample mean (8) shows that population mean is smaller than 10. So we take the opposite as the null hypothesis: H_0: $\mu \geq 10$.

So we have this sampling distribution about the mean:

$$\overline{x} \sim N\left(10, \frac{4}{\sqrt{9}}\right) : N(10, 1.33)$$

We will solve it by p-value, but it may be solved also by the critical region.

We reject the null hypothesis when $\overline{x}$ is very small. So:

$$P[\overline{x} < 8] = P\Big[Z < \frac{8-10}{1.33}\Big] = P[Z < -2] = 0.03$$

p-value (0.03) is bigger than α, so we accept H_0 and state that the evidence against the null hypothesis, in order to state that the specification is not met, is not strong enough.

A machine's output is 40 units on average, and allegedly follows a normal distribution. In order to have independent data, we have drawn 4 output data from different days: 38-39-35-36. What should we decide with those data about the average output? Significance level: 5%.

The don't state any null hypothesis, and there isn't any claim or question about the population mena. So, following the [3] criterion we look at the sample mean to give H_0:

$$\overline{x} = 37$$

It looks like the population mean has decreased ($37 < 40$), so we take the opposite as H_0:

$$H_0 : \mu > 40$$

Population is normal and σ is unknown. So we must calculate the t statistic, that follows a t_{4-1} distribution. Because of the format of the table, we cannot give the p-value and use the critical region method.

We reject H_0 (big μ values), when $\overline{x}$ and also its t statistics are small. So, test is one-sided or one-tailed and critical region is on the lower side.

Concretely,for a t_{4-1} distribution, the value that leaves below it a 5% probability is -2.35. Below to it, we would reject H_0.

We calculate the corresponding t statistic:

$$\overline{x} = 37$$

$$\hat{s} = \sqrt{\frac{(38-37)^2 + (39-37)^2 + (35-37)^2 + (36-37)^2}{4-1}} = 1.82$$

$$t = \frac{\overline{x} - \mu}{\frac{\hat{s}}{\sqrt{n}}} = \frac{37-40}{\frac{1.82}{\sqrt{4}}} = -3.29$$

Int's below the critical value, so we reject H_0. So we state that mean production has decreased.

If we want to solve it not by t, but by x:

$$t_{4-1,0.95} = \frac{\overline{x}_0 - 40}{\frac{1.82}{\sqrt{4}}} = -2.35 \rightarrow \overline{x} = 37.86$$

Sample mean is under the critical value for the sample mean. So we reject H_0.

We have drawn 10 data from an alleged normal distribution. Sample variance (without correction) is 36. Test whether the standard deviation in the population is 5 or less. Significance level: 0.01.

$H_0 : \sigma^2 \leq 25$ (as directly stated in the problem) ([1] criterion)

We reject that hypothesis when its estimator (s^2) and also its corresponding $\frac{ns^2}{\sigma^2}$ is big enough. So the critical region is on the upper side and will have a 0.01 probability.

$\frac{ns^2}{\sigma^2}$ follows a chi-square distribution with 10-1=9 degrees of freedom. In that distribution, the critical value will leave above it (rejection on the upper side) a 0.01 probability, and below it 0.99. So, looking into the tables, we have as critical value 21.7.

The value for the statistic is:

$$\frac{ns^2}{\sigma^2} = \frac{10 \times 36}{25} = 14.4$$

As it's bigger than the critical value, it's on the acceptance region ans so we will accept H_0.

Remark: here we can't apply p-value method, as tables are given only for some probabilities.